GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 23, 2003, 12:45:21 ; Search time 124.667 Seconds
(without alignments)
433.540 Million cell updates/sec

Title: US-09-700-148-15
Perfect score: 24
Sequence: 1 gtgaaattatcgccacgttcgggc 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

cal number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	24	100.0	24	21	AAZ43972	Salmonella sp. det
	2	24	100.0	26	14	AAQ34560	Sequence of probe/
c	3	24	100.0	288	23	AAS51162	Salmonella typhimu
	4	24	100.0	310	21	AAZ43961	Salmonella sp. det
c	5	24	100.0	405	23	AAS51146	Salmonella typhimu
c	6	24	100.0	405	23	AAS51149	Salmonella typhimu
c	7	24	100.0	440	23	AAS51168	Salmonella typhimu
	8	24	100.0	2058	14	AAQ34562	Sequence of the in
c	9	17.8	74.2	1476	21	AAZ54345	Neisseria gonorrhe
c	10	17.8	74.2	49646	21	AAA81457	N. meningitidis pa
	11	17.8	74.2	349980	21	AAF21608	Neisseria meningit
	12	17.4	72.5	230	22	AAH36436	Human colon cancer
c	13	17.2	71.7	1477	21	AAC39037	Arabidopsis thalia
c	14	16.8	70.0	3078	23	AAS93759	DNA encoding novel
	15	16.8	70.0	7237	20	AAX13176	Enterococcus faeca
	16	16.8	70.0	24789	23	ABL28640	Drosophila melanog
	17	16.6	69.2	1171	22	AAH49321	V. vinifera aquapo
c	18	16.6	69.2	3664	15	AAQ54459	NaeI restriction e
c	19	16.6	69.2	4205	23	ABL21992	Drosophila melanog
	20	16.2	67.5	211	22	AAH81223	Escherichia coli n
	21	16.2	67.5	582	24	ABN64665	Human cancer relat
c	22	16.2	67.5	719	21	AAF12228	Aspergillus oryzae
	23	16.2	67.5	1334	21	AAA39409	Rice SYR2 homologu
c	24	16.2	67.5	1473	21	AAZ54346	Neisseria meningit
c	25	16.2	67.5	1473	21	AAZ54347	Neisseria meningit
c	26	16.2	67.5	1564	23	ABL14907	Drosophila melanog
c	27	16.2	67.5	1923	24	ABQ70655	Listeria monocytog
	28	16.2	67.5	1944	23	AAS53484	Haemophilus influe
	29	16.2	67.5	3162	23	ABL04446	Drosophila melanog
	30	16.2	67.5	3633	23	ABL14906	Drosophila melanog
	31	16.2	67.5	5626	23	ABL14766	Drosophila melanog
c	32	16.2	67.5	21091	21	AAA81523	N. meningitidis pa
	33	16.2	67.5	28098	23	ABL14892	Drosophila melanog
c	34	16.2	67.5	349980	21	AAF21544	Neisseria meningit
c	35	16.2	67.5	349980	21	AAF21607	Neisseria meningit
	36	16	66.7	306	15	AAQ57756	E.coli F-plasmid c
	37	16	66.7	670	24	AAD27074	PCR fragment used
c	38	16	66.7	1560	20	AAX38288	E. coli nrdB DNA.
c	39	16	66.7	1806	23	AAS74533	DNA encoding novel
c	40	16	66.7	1806	23	AAS94226	DNA encoding novel
c	41	16	66.7	2460	23	ABL03553	Drosophila melanog
	42	16	66.7	2717	21	AAC55422	Entry vector pENTR
	43	16	66.7	2717	21	AAC55437	Entry vector pENTR
	44	16	66.7	2718	21	AAC55425	Entry vector pENTR
	45	16	66.7	2720	21	AAC55431	Entry vector pENTR

ALIGNMENTS

RESULT 1
AAZ43972
ID AAZ43972 standard; DNA; 24 BP.
XX
AC AAZ43972;
XX
DT 17-MAR-2000 (first entry)
XX
DE Salmonella sp. detecting primer #269*.
XX
KW Detection; microorganism; primer; probe; cosmetic; food; ss.
XX
OS Salmonella sp.
XX
PN WO9958713-A2.
XX
PD 18-NOV-1999.
XX
PF 10-MAY-1999; 99WO-DE01471.
XX
PR 12-MAY-1998; 98DE-1022108.
XX
PA (BIOI-) BIOINSIDE GMBH.
XX
PI Gerbling K, Lauter F, Grohmann L;
XX
DR WPI; 2000-072341/06.
XX
PT A test kit for detecting microbially soiled, non sterile products,
PT especially pharmaceuticals and cosmetics -
XX
PS Claim 1g(iv); Page 71; 77pp; German.

```
XX
CC   This invention describes a novel test kit to detect microbially soiled,
CC   non-sterile products, in particular after GMP-rich lines, also in
CC   cosmetics and food. The method involves the use of DNA fragment having a
CC   forward primer, probe, a reverse primer and if necessary a spacer
CC   oligonucleotide. The test kit and method are useful for economic
CC   detection of germs in pharmaceutical and cosmetic products.  In
CC   particular the method is useful for detecting E. coli, P. aeruginosa,
CC   S. aureus and Salmonella.
XX
SQ   Sequence 24 BP; 5 A; 6 C; 7 G; 6 T; 0 other;

  Query Match             100.0%;  Score 24;  DB 21;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 0.0054;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTGAAATTATCGCCACGTTCGGGC 24
           ||||||||||||||||||||||||
Db       1 GTGAAATTATCGCCACGTTCGGGC 24


  SULT 2
 .AQ34560
ID   AAQ34560 standard; DNA; 26 BP.
XX
AC   AAQ34560;
XX
DT   05-JUL-1993  (first entry)
XX
DE   Sequence of probe/primer for the invA gene in Salmonella.
XX
KW   Inv gene; invasion gene; probe; primer; detection; diagnosis; ss.
XX
OS   Synthetic.
XX
PN   WO9304202-A.
XX
PD   04-MAR-1993.
XX
PF   19-AUG-1992;   92WO-US06984.
XX
PR   22-AUG-1991;   91US-0749447.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Curtiss R,  Galan J;
XX
DR   WPI; 1993-094027/11.
XX
     New polynucleotide probes and primers from Salmonella inv gene -
     for detecting Salmonella nucleotide sequences in a sample
XX
PS   Claim 3; Page 45; 62pp; English.
XX
CC   Four genes are involved in the invasive phenotype of S.typhimurium
CC   strain DB4673. These are invA, invB, invC, and InvD, encoding
CC   proteins of 54,64,47 and 30kDa respectively. invA,invB and invC are
CC   in the same transcriptional unit. The invA gene was sequenced
CC   (AAQ34562) and based on this sequence, primers were synthesised using
CC   std. techniques. The primers consisted of the sequences in AAQ34560
CC   and AAQ34561. These primers were used in PCR amplification studies of
CC   636 strains of Salmonella belonging to over 100 serotypes. Of the
CC   636 straines tested, 634 were specifically detected and no non-
CC   Salmonella strains were specifically amplified.
XX
SQ   Sequence 26 BP; 7 A; 6 C; 7 G; 6 T; 0 other;

  Query Match             100.0%;  Score 24;  DB 14;  Length 26;
  Best Local Similarity   100.0%;  Pred. No. 0.0054;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTGAAATTATCGCCACGTTCGGGC 24
           ||||||||||||||||||||||||
Db       1 GTGAAATTATCGCCACGTTCGGGC 24


RESULT 3
AAS51162/c
ID   AAS51162 standard; DNA; 288 BP.
XX
AC   AAS51162;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Salmonella typhimurium cellular proliferation inhibitory sequence #60.
XX
KW   Antisense; ss; prokaryotic cellular proliferation;
KW   antibiotic; antibacterial; drug design.
XX
OS   Salmonella typhimurium.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Claim 1; Seq ID No 3739; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence is an antisense
CC   oligonucleotide of the invention.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 288 BP; 99 A; 71 C; 65 G; 53 T; 0 other;

  Query Match             100.0%;  Score 24;  DB 23;  Length 288;
  Best Local Similarity   100.0%;  Pred. No. 0.0077;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTGAAATTATCGCCACGTTCGGGC 24
           ||||||||||||||||||||||||
Db     156 GTGAAATTATCGCCACGTTCGGGC 133
```

```
RESULT 4
AAZ43961
ID   AAZ43961 standard; DNA; 310 BP.
XX
AC   AAZ43961;
XX
DT   17-MAR-2000  (first entry)
XX
DE   Salmonella sp. detecting primer #1.
XX
KW   Detection; microorganism; primer; probe; cosmetic; food; ss.
XX
OS   Salmonella sp.
XX
PN   WO9958713-A2.
XX
PD   18-NOV-1999.

     10-MAY-1999;   99WO-DE01471.
XX
PR   12-MAY-1998;   98DE-1022108.
XX
PA   (BIOI-) BIOINSIDE GMBH.
XX
PI   Gerbling K,  Lauter F,  Grohmann L;
XX
DR   WPI; 2000-072341/06.
XX
PT   A test kit for detecting microbially soiled, non sterile products,
PT   especially pharmaceuticals and cosmetics  -
XX
PS   Example 24; Page 69; 77pp; German.
XX
CC   This invention describes a novel test kit to detect microbially soiled,
CC   non-sterile products, in particular after GMP-rich lines, also in
CC   cosmetics and food. The method involves the use of DNA fragment having a
CC   forward primer, probe, a reverse primer and if necessary a spacer
CC   oligonucleotide. The test kit and method are useful for economic
CC   detection of germs in pharmaceutical and cosmetic products.  In
CC   particular the method is useful for detecting E. coli, P. aeruginosa,
CC   S. aureus and Salmonella.
XX
SQ   Sequence 310 BP; 64 A; 63 C; 94 G; 89 T; 0 other;

  Query Match             100.0%;  Score 24;  DB 21;  Length 310;
  Best Local Similarity   100.0%;  Pred. No. 0.0078;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTGAAATTATCGCCACGTTCGGGC 24
           ||||||||||||||||||||||||
Db      19 GTGAAATTATCGCCACGTTCGGGC 42


RESULT 5
AAS51146/c
ID   AAS51146 standard; DNA; 405 BP.
XX
AC   AAS51146;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Salmonella typhimurium cellular proliferation inhibitory sequence #44.
XX
KW   Antisense; ss; prokaryotic cellular proliferation;
KW   antibiotic; antibacterial; drug design.
XX
OS   Salmonella typhimurium.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Claim 1; Seq ID No 3723; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence is an antisense
CC   oligonucleotide of the invention.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 405 BP; 141 A; 82 C; 86 G; 96 T; 0 other;

  Query Match             100.0%;  Score 24;  DB 23;  Length 405;
  Best Local Similarity   100.0%;  Pred. No. 0.0081;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTGAAATTATCGCCACGTTCGGGC 24
           ||||||||||||||||||||||||
Db      43 GTGAAATTATCGCCACGTTCGGGC 20


RESULT 6
AAS51149/c
ID   AAS51149 standard; DNA; 405 BP.
XX
AC   AAS51149;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Salmonella typhimurium cellular proliferation inhibitory sequence #47.
XX
KW   Antisense; ss; prokaryotic cellular proliferation;
KW   antibiotic; antibacterial; drug design.
XX
OS   Salmonella typhimurium.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
```

```
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Claim 1; Seq ID No 3726; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence is an antisense
CC   oligonucleotide of the invention.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 405 BP; 141 A; 82 C; 86 G; 96 T; 0 other;

  Query Match             100.0%;  Score 24;  DB 23;  Length 405;
  Best Local Similarity   100.0%;  Pred. No. 0.0081;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTGAAATTATCGCCACGTTCGGGC 24
           ||||||||||||||||||||||||
Db      43 GTGAAATTATCGCCACGTTCGGGC 20


RESULT 7
AAS51168/c
ID   AAS51168 standard; DNA; 440 BP.
XX
AC   AAS51168;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Salmonella typhimurium cellular proliferation inhibitory sequence #66.
XX
KW   Antisense; ss; prokaryotic cellular proliferation;
KW   antibiotic; antibacterial; drug design.
XX
OS   Salmonella typhimurium.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Claim 1; Seq ID No 3745; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence is an antisense
CC   oligonucleotide of the invention.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 440 BP; 158 A; 92 C; 93 G; 97 T; 0 other;

  Query Match             100.0%;  Score 24;  DB 23;  Length 440;
  Best Local Similarity   100.0%;  Pred. No. 0.0082;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTGAAATTATCGCCACGTTCGGGC 24
           ||||||||||||||||||||||||
Db      74 GTGAAATTATCGCCACGTTCGGGC 51


RESULT 8
AAQ34562
ID   AAQ34562 standard; DNA; 2058 BP.
XX
AC   AAQ34562;
XX
DT   05-JUL-1993  (first entry)
XX
DE   Sequence of the invA gene in Salmonella.
XX
KW   Inv gene; invasion gene; probe; primer; detection; diagnosis; ss.
XX
OS   Synthetic.
XX
PN   WO9304202-A.
XX
PD   04-MAR-1993.
XX
PF   19-AUG-1992;   92WO-US06984.
XX
PR   22-AUG-1991;   91US-0749447.
```

```
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Curtiss R,  Galan J;
XX
DR   WPI; 1993-094027/11.
XX
PT   New polynucleotide probes and primers from Salmonella inv gene -
PT   for detecting Salmonella nucleotide sequences in a sample
XX
PS   Example; Fig 1; 62pp; English.
XX
CC   Four genes are involved in the invasive phenotype of S.typhimurium
CC   strain DB4673. These are invA, invB, invC, and InvD, encoding
CC   proteins of 54,64,47 and 30kDa respectively. invA,invB and invC are
CC   in the same transcriptional unit. The invA gene was sequenced
CC   (AAQ34562) and based on this sequence, primers were synthesised using
CC   std. techniques. The primers consisted of the sequences in AAQ34560
CC   and AAQ34561. These primers were used in PCR amplification studies of
     636 strains of Salmonella belonging to over 100 serotypes. Of the
     636 straines tested, 634 were specifically detected and no non-
CC   Salmonella strains were specifically amplified.
XX
SQ   Sequence 2058 BP; 493 A; 414 C; 523 G; 628 T; 0 other;

  Query Match             100.0%;  Score 24;  DB 14;  Length 2058;
  Best Local Similarity   100.0%;  Pred. No. 0.01;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTGAAATTATCGCCACGTTCGGGC 24
           ||||||||||||||||||||||||
Db     287 GTGAAATTATCGCCACGTTCGGGC 310


RESULT 9
AAZ54345/c
ID   AAZ54345 standard; DNA; 1476 BP.
XX
AC   AAZ54345;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria gonorrheae ORF 769 partial DNA sequence SEQ ID NO:2639.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy; ds.
XX
OS   Neisseria gonorrheae.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   P-PSDB; AAY75583.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 7; Page 1250; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence 1476 BP; 354 A; 368 C; 446 G; 308 T; 0 other;

  Query Match             74.2%;  Score 17.8;  DB 21;  Length 1476;
  Best Local Similarity   90.5%;  Pred. No. 18;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 TGAAATTATCGCCACGTTCGG 22
           ||||||| |||||||| ||||
Db    1177 TGAAATTGTCGCCACGGTCGG 1157


RESULT 10
AAA81457/c
ID   AAA81457 standard; DNA; 49646 BP.
XX
AC   AAA81457;
XX
DT   04-DEC-2000  (first entry)
XX
DE   N. meningitidis partial DNA sequence gnm_5 SEQ ID NO:5.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; genome; immunogenic;
KW   antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW   Meningococcus B; MenB; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200022430-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-US23573.
XX
PR   09-OCT-1998;   98US-0103794.
PR   30-APR-1999;   99US-0132068.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Frazer CM,  Hickey E,  Peterson J,  Tettelin H,  Venter JC;
PI   Masignani V,  Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V;
PI   Rappuoli R,  Pizza M;
XX
DR   WPI; 2000-318079/27.
XX
PT   Isolated nucleotide sequences of Neisseria meningitidis which can be
PT   used in the diagnosis and treatment of N. meningitidis infection and
PT   other Neisserial infections, for example, N.gonorrhoea -
XX
PS   Claim 7; Page 274-288; 1760pp; English.
XX
CC   The present invention describes methods of obtaining immunogenic
CC   proteins from Neisseria genomic sequences. AAA81453 to AAA82414
```

CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to
CC Neisserial bacteria. For example, some of the identified proteins could
CC be components of vaccines against Meningococcus B; against all serotypes;
CC and/or against all pathogenic Neissariae. Identification of sequences
CC from the bacterium will also facilitate production of biological probes,
CC particularly organism-specific probes. Attempts to make efficacious
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC Multivalent vaccines have also been tried but none have successfully
CC overcome antigenic variability. The provision of further, complete
CC sequences may provide an opportunity to identify secreted or surface
CC exposed proteins that may be presumed targets for the immune system and
CC which are not antigenically variable or at least more conserved than
CC other more variable regions.
XX
SQ Sequence 49646 BP; 12331 A; 14486 C; 11862 G; 10966 T; 1 other;

Query Match 74.2%; Score 17.8; DB 21; Length 49646;
Best Local Similarity 90.5%; Pred. No. 31;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 GAAATTATCGCCACGTTCGGG 23
||||| |||||||| ||||||
Db 4905 GAAATCATCGCCACTTTCGGG 4885

RESULT 11
AAF21608
ID AAF21608 standard; DNA; 349980 BP.
XX
AC AAF21608;
XX
DT 13-MAR-2001 (first entry)
XX
DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:109.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; immunogenic; vaccine;
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW ds.
XX
OS Neisseria meningitidis.
XX
PN WO200066791-A1.
XX
PD 09-NOV-2000.
XX
PF 08-MAR-2000; 2000WO-US05928.
XX
PR 30-APR-1999; 99US-0132068.
PR 08-OCT-1999; 99WO-US23573.
PR 28-FEB-2000; 2000GB-0004695.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
PI Frazer CM, Grandi G;
XX
DR WPI; 2000-647603/62.
XX
PT Neisseria meningitidis B full length genome sequence and open reading
PT frames are used to detect, treat and prevent Neisserial infections -
XX
PS Claim 7; Appendix A; 692pp; English.
XX
CC The present invention describes the full length genome of
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC to AAF21613 represent fragments of the NMB genomic sequence, as the
CC sequence was too long to go in a record on its own it was split into 8
CC sequences which overlap each other at the beginning and end of each
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC AAF21606 represent PCR primers which are used in the exemplification of
CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC presence of Neisserial bacteria or of antibodies raised to Neisserial
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.
XX
SQ Sequence 349980 BP; 82523 A; 82940 C; 96712 G; 87805 T; 0 other;

Query Match 74.2%; Score 17.8; DB 21; Length 349980;
Best Local Similarity 90.5%; Pred. No. 41;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 GAAATTATCGCCACGTTCGGG 23
||||| |||||||| ||||||
Db 64028 GAAATCATCGCCACTTTCGGG 64048

RESULT 12
AAH36436
ID AAH36436 standard; cDNA; 230 BP.
XX
AC AAH36436;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:3518.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI; 2001-235357/24.
DR P-PSDB; AAG77031.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 1; Page 5307-5308; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon

```
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where
CC   the proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene
CC   therapy and vaccine production. N and P may be used in the prevention,
CC   diagnosis and treatment of diseases associated with inappropriate P
CC   expression. For example, N and P may be used to treat disorders
CC   associated with decreased expression by rectifying mutations or deletions
CC   in a patient's genome that affect the activity of P by expressing
CC   inactive proteins or to supplement the patients own production of P.
CC   Additionally, N may be used to produce the colon cancer-associated Ps,
CC   by inserting the nucleic acids into a host cell and culturing the cell
CC   to express the proteins. N and P can be used in the prevention, diagnosis
CC   and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC   and AAB77789 represent sequences used in the exemplification of the
CC   present invention.
CC   N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC   missing at time of publication, meaning no sequences are present for
CC   SEQ ID NO:1027 to 1052, 7921 and 7922.
XX

     Sequence 230 BP; 51 A; 50 C; 71 G; 47 T; 11 other;

  Query Match             72.5%;  Score 17.4;  DB 22;  Length 230;
  Best Local Similarity   90.0%;  Pred. No. 23;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GTGAAATTATCGCCACGTTC 20
           || |||||| ||||||||||
Db      13 GTNAAATTACCGCCACGTTC 32


RESULT 13
AAC39037/c
ID   AAC39037 standard; DNA; 1477 BP.
XX
AC   AAC39037;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 23132.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
```

PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
10-AUG-1999; 99US-0148171.
11-AUG-1999; 99US-0148319.
12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
23-SEP-1999; 99US-0155486.
24-SEP-1999; 99US-0155659.
28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 71.7%; Score 17.2; DB 21; Length 1477;
Best Local Similarity 86.4%; Pred. No. 38;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GTGAAATTATCGCCACGTTCGG 22
||||||||||||||| ||||
Db 190 GTGAAATTATCGCCAATATCGG 169

RESULT 14
AAS93759/c
ID AAS93759 standard; cDNA; 3078 BP.
XX
AC AAS93759;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #29563.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG29572.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 29563; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.

```
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3078 BP; 632 A; 859 C; 866 G; 721 T; 0 other;

  Query Match             70.0%;  Score 16.8;  DB 23;  Length 3078;
  Best Local Similarity   90.0%;  Pred. No. 69;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 TGAAATTATCGCCACGTTCG 21
           |||||||| |||| ||||||
Db    2995 TGAAATTAACGCCGCGTTCG 2976


RESULT 15
AAX13176
ID   AAX13176 standard; DNA; 7237 BP.
XX
AC   AAX13176;
XX
DT   19-MAR-1999  (first entry)
XX
DE   Enterococcus faecalis genome contig SEQ ID NO:239.
XX
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850555-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US08985.
XX
PR   14-NOV-1997;   97US-0066009.
PR   06-MAY-1997;   97US-0044031.
PR   16-MAY-1997;   97US-0046655.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.

     Barash SC,  Dillon PJ,  Kunsch CA;
XX
DR   WPI; 1999-045171/04.
XX
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides
PT   - used to develop products for the detection of Enterococcus and for
PT   use in vaccines for prevention or attenuation of Enterococcus
PT   infection.
XX
PS   Claim 1; Page 1196-1200; 2084pp; English.
XX
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   AAX12938 to AAX13919 represent these nucleotide sequences which are
CC   primary nucleotide sequences, also known as contigs. The computer-based
CC   system can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence
CC   of Enterococcus faecalis in samples. They can also be used for
CC   diagnosing Enterococcal infection in an animal and monitoring
CC   progression of disease, and for identifying agents which can be used to
CC   modulate the growth or pathogenicity of Enterococcus faecalis, or
CC   another related organism, in vivo or in vitro. In particular the
CC   polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC   can be used in vaccines to prevent or attenuate an Enterococcal
CC   infection.
XX
SQ   Sequence 7237 BP; 2361 A; 1134 C; 1596 G; 2138 T; 8 other;

  Query Match             70.0%;  Score 16.8;  DB 20;  Length 7237;
  Best Local Similarity   90.0%;  Pred. No. 78;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GTGAAATTATCGCCACGTTC 20
           ||  ||||||||||||||||
Db    2089 GTCGAATTATCGCCACGTTC 2108


Search completed: January 23, 2003, 13:54:54
Job time : 159.667 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 23, 2003, 12:45:21 ; Search time 124.667 Seconds
(without alignments)
433.540 Million cell updates/sec

Title: US-09-700-148-16
Perfect score: 24
Sequence: 1 cttctctattgtcaccgtggtcca 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	24	100.0	24	21	AAZ43973	Salmonella sp. det
c	2	24	100.0	288	23	AAS51162	Salmonella typhimu
	3	24	100.0	310	21	AAZ43961	Salmonella sp. det
	4	24	100.0	2058	14	AAQ34562	Sequence of the in
c	5	17.6	73.3	595	21	AAF07983	Fusarium venenatum
	6	17.6	73.3	1385	22	AAS02654	Human secreted pro
c	7	17.2	71.7	1638	17	AAT06480	Cystathionine gamm
c	8	17.2	71.7	1969	16	AAQ98751	DNA encoding murin
c	9	17.2	71.7	1992	24	ABK63809	Rat sequence diffe
	10	17.2	71.7	2702	23	ABL27042	Drosophila melanog
c	11	17.2	71.7	3639	17	AAT06481	Cystathionine gamm
c	12	16.8	70.0	261	19	AAV15589	Human HPK-1A C21.7
	13	16.8	70.0	400	22	AAF66319	Novel human polynu
c	14	16.8	70.0	502	24	ABQ55843	Human ovarian anti
	15	16.8	70.0	904	22	AAS33242	DNA encoding human
	16	16.8	70.0	904	22	AAS34994	cDNA encoding nove
c	17	16.8	70.0	943	22	AAH44805	Murine cDNA encodi
c	18	16.8	70.0	1026	19	AAV64572	Human Fanconi anae
	19	16.8	70.0	1268	22	AAS34864	cDNA encoding nove
c	20	16.8	70.0	1373	24	ABK92263	Prostate cancer-as
	21	16.8	70.0	1725	22	AAI59520	Human polynucleoti
	22	16.8	70.0	1900	22	AAF27662	DNA encoding human
	23	16.8	70.0	2004	23	AAS91802	DNA encoding novel
	24	16.8	70.0	8880	24	AAS96692	Arabidopsis DMT2 (
	25	16.8	70.0	9192	22	AAS33461	DNA encoding human
	26	16.8	70.0	25715	22	AAS33462	DNA encoding human
c	27	16.6	69.2	63	22	AAC86151	Forward primer Spe
c	28	16.6	69.2	462	22	AAI81477	Human polynucleoti
c	29	16.6	69.2	684	24	ABL01494	Murine apoptosis r
c	30	16.6	69.2	1464	21	AAZ61782	cDNA encoding muri
c	31	16.6	69.2	1464	22	AAC99715	Skin cell cDNA, SE
c	32	16.6	69.2	1464	24	ABL34867	Murine cDNA isolat
c	33	16.6	69.2	1627	22	AAC99790	Skin cell cDNA, SE
c	34	16.6	69.2	1627	24	ABL34942	Murine cDNA isolat
c	35	16.6	69.2	1633	24	ABL34763	Murine cDNA isolat
c	36	16.6	69.2	1635	21	AAZ61678	cDNA encoding muri
c	37	16.6	69.2	1635	22	AAC99611	Skin cell cDNA, SE
c	38	16.6	69.2	2140	19	AAV42316	Macadamia integrif
c	39	16.6	69.2	2171	19	AAV42311	Macadamia integrif
	40	16.6	69.2	4858	23	ABL18598	Drosophila melanog
	41	16.6	69.2	8560	23	ABL02936	Drosophila melanog
c	42	16.6	69.2	10382	22	AAK67484	Human immune/haema
c	43	16.6	69.2	19191	22	AAK67485	Human immune/haema
c	44	16.6	69.2	2365589	24	ABA90521	Genomic sequence o
	45	16.2	67.5	1245	21	AAA05515	Streptococcus pneu

ALIGNMENTS

RESULT 1
AAZ43973
ID AAZ43973 standard; DNA; 24 BP.
XX
AC AAZ43973;
XX
DT 17-MAR-2000 (first entry)
XX
DE Salmonella sp. detecting probe #333*.
XX
KW Detection; microorganism; primer; probe; cosmetic; food; ss.
XX
OS Salmonella sp.
XX
PN WO9958713-A2.
XX
PD 18-NOV-1999.
XX
PF 10-MAY-1999; 99WO-DE01471.
XX
PR 12-MAY-1998; 98DE-1022108.
XX
PA (BIOI-) BIOINSIDE GMBH.
XX
PI Gerbling K, Lauter F, Grohmann L;
XX
DR WPI; 2000-072341/06.
XX
PT A test kit for detecting microbially soiled, non sterile products,
PT especially pharmaceuticals and cosmetics -
XX
PS Claim 1g(iv); Page 71; 77pp; German.

```
XX
CC   This invention describes a novel test kit to detect microbially soiled,
CC   non-sterile products, in particular after GMP-rich lines, also in
CC   cosmetics and food. The method involves the use of DNA fragment having a
CC   forward primer, probe, a reverse primer and if necessary a spacer
CC   oligonucleotide. The test kit and method are useful for economic
CC   detection of germs in pharmaceutical and cosmetic products.  In
CC   particular the method is useful for detecting E. coli, P. aeruginosa,
CC   S. aureus and Salmonella.
XX
SQ   Sequence 24 BP; 3 A; 8 C; 4 G; 9 T; 0 other;

  Query Match            100.0%;  Score 24;  DB 21;  Length 24;
  Best Local Similarity  100.0%;  Pred. No. 0.046;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CTTCTCTATTGTCACCGTGGTCCA 24
           ||||||||||||||||||||||||
Db       1 CTTCTCTATTGTCACCGTGGTCCA 24


  SULT 2
  AS51162/c
ID   AAS51162 standard; DNA; 288 BP.
XX
AC   AAS51162;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Salmonella typhimurium cellular proliferation inhibitory sequence #60.
XX
KW   Antisense; ss; prokaryotic cellular proliferation;
KW   antibiotic; antibacterial; drug design.
XX
OS   Salmonella typhimurium.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.

     (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Claim 1; Seq ID No 3739; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence is an antisense
CC   oligonucleotide of the invention.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 288 BP; 99 A; 71 C; 65 G; 53 T; 0 other;

  Query Match            100.0%;  Score 24;  DB 23;  Length 288;
  Best Local Similarity  100.0%;  Pred. No. 0.066;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CTTCTCTATTGTCACCGTGGTCCA 24
           ||||||||||||||||||||||||
Db      92 CTTCTCTATTGTCACCGTGGTCCA 69


RESULT 3
AAZ43961
ID   AAZ43961 standard; DNA; 310 BP.
XX
AC   AAZ43961;
XX
DT   17-MAR-2000  (first entry)
XX
DE   Salmonella sp. detecting primer #1.
XX
KW   Detection; microorganism; primer; probe; cosmetic; food; ss.
XX
OS   Salmonella sp.
XX
PN   WO9958713-A2.
XX
PD   18-NOV-1999.
XX
PF   10-MAY-1999;   99WO-DE01471.
XX
PR   12-MAY-1998;   98DE-1022108.
XX
PA   (BIOI-) BIOINSIDE GMBH.
XX
PI   Gerbling K,  Lauter F,  Grohmann L;
XX
DR   WPI; 2000-072341/06.
XX
PT   A test kit for detecting microbially soiled, non sterile products,
PT   especially pharmaceuticals and cosmetics -
XX
PS   Example 24; Page 69; 77pp; German.
XX
CC   This invention describes a novel test kit to detect microbially soiled,
CC   non-sterile products, in particular after GMP-rich lines, also in
CC   cosmetics and food. The method involves the use of DNA fragment having a
CC   forward primer, probe, a reverse primer and if necessary a spacer
CC   oligonucleotide. The test kit and method are useful for economic
CC   detection of germs in pharmaceutical and cosmetic products.  In
CC   particular the method is useful for detecting E. coli, P. aeruginosa,
CC   S. aureus and Salmonella.
XX
SQ   Sequence 310 BP; 64 A; 63 C; 94 G; 89 T; 0 other;

  Query Match            100.0%;  Score 24;  DB 21;  Length 310;
  Best Local Similarity  100.0%;  Pred. No. 0.067;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CTTCTCTATTGTCACCGTGGTCCA 24
           ||||||||||||||||||||||||
Db      83 CTTCTCTATTGTCACCGTGGTCCA 106
```

```
RESULT 4
AAQ34562
ID   AAQ34562 standard; DNA; 2058 BP.
XX
AC   AAQ34562;
XX
DT   05-JUL-1993  (first entry)
XX
DE   Sequence of the invA gene in Salmonella.
XX
KW   Inv gene; invasion gene; probe; primer; detection; diagnosis; ss.
XX
OS   Synthetic.
XX
PN   WO9304202-A.
XX
PD   04-MAR-1993.
XX
PF   19-AUG-1992;   92WO-US06984.

     22-AUG-1991;   91US-0749447.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Curtiss R,  Galan J;
XX
DR   WPI; 1993-094027/11.
XX
PT   New polynucleotide probes and primers from Salmonella inv gene -
PT   for detecting Salmonella nucleotide sequences in a sample
XX
PS   Example; Fig 1; 62pp; English.
XX
CC   Four genes are involved in the invasive phenotype of S.typhimurium
CC   strain DB4673. These are invA, invB, invC, and InvD, encoding
CC   proteins of 54,64,47 and 30kDa respectively. invA,invB and invC are
CC   in the same transcriptional unit. The invA gene was sequenced
CC   (AAQ34562) and based on this sequence, primers were synthesised using
CC   std. techniques. The primers consisted of the sequences in AAQ34560
CC   and AAQ34561. These primers were used in PCR amplification studies of
CC   636 strains of Salmonella belonging to over 100 serotypes. Of the
CC   636 straines tested, 634 were specifically detected and no non-
CC   Salmonella strains were specifically amplified.
XX
SQ   Sequence 2058 BP; 493 A; 414 C; 523 G; 628 T; 0 other;

  Query Match             100.0%;  Score 24;  DB 14;  Length 2058;
  Best Local Similarity   100.0%;  Pred. No. 0.089;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CTTCTCTATTGTCACCGTGGTCCA 24
           ||||||||||||||||||||||||
Db     351 CTTCTCTATTGTCACCGTGGTCCA 374


RESULT 5
AAF07983/c
ID   AAF07983 standard; cDNA; 595 BP.
XX
AC   AAF07983;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Fusarium venenatum EST SEQ ID NO:506.
XX
KW   Multiple gene expression; filamentous fungal cell; EST;
KW   expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW   Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW   culture condition; environmental stress; spore morphogenesis;
KW   metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS   Fusarium venenatum.
XX
PN   WO200056762-A2.
XX
PD   28-SEP-2000.
XX
PF   22-MAR-2000; 2000WO-US07781.
XX
PR   22-MAR-1999;   99US-0273623.
XX
PA   (NOVO ) NOVO NORDISK BIOTECH INC.
PA   (NOVO ) NOVO NORDISK AS.
XX
PI   Berka RM,  Rey MW,  Shuster JR,  Kauppinen S,  Clausen IG,  Olsen PB;
XX
DR   WPI; 2000-594572/56.
XX
PT   Monitoring differential expression of genes in filamentous fungal cells
PT   uses fluorescence-labeled nucleic acids isolated from the cells and a
PT   substrate of expressed sequence tags -
XX
PS   Claim 86; Page 579; 3161pp; English.
XX
CC   The present invention describes a method for monitoring differential
CC   expression of genes in a first filamentous fungal (FF) cell relative to
CC   expression of the same genes in one or more second filamentous fungal
CC   cells. The method uses fluorescence-labeled nucleic acids isolated from
CC   the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC   are used in the methods for monitoring differential expression of genes
CC   in a first filamentous fungal (FF) cell relative to expression of the
CC   same genes in one or more second filamentous fungal cells. Monitoring
CC   the global expression of genes from FF cells allows the production
CC   potential of the microorganisms to be improved. New genes may be
CC   discovered, possible functions of unknown open reading frames can be
CC   identified and gene copy number variation and stability can be
CC   monitored. The expression of genes can be used to study how FF cells
CC   adapt to changes in culture conditions, environmental stress, spore
CC   morphogenesis, recombination, metabolic or catabolic pathway
CC   engineering. Using ESTs provides several advantages over genomic or
CC   random cDNA clones including elimination of redundancy as one spot on an
CC   array equals one gene or open reading frame, and organisation of the
CC   microarrays based on function of the gene products to facilitate
CC   analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC   Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC   niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC   AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC   all specifically claimed in the present invention.
XX
SQ   Sequence 595 BP; 142 A; 149 C; 152 G; 146 T; 6 other;

  Query Match             73.3%;  Score 17.6;  DB 21;  Length 595;
  Best Local Similarity   83.3%;  Pred. No. 86;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 CTTCTCTATTGTCACCGTGGTCCA 24
           |||||| | ||||||| ||| |||
Db     428 CTTCTCCAGTGTCACCTTGGGCCA 405


RESULT 6
AAS02654
ID   AAS02654 standard; cDNA; 1385 BP.
XX
AC   AAS02654;
XX
DT   18-JUL-2001  (first entry)
XX
DE   Human secreted protein gene #15.
XX
KW   Human secreted protein; autoimmune disorder; hyperproliferative disorder;
KW   cardiovascular disorder; cerebrovascular disorder; angiogenesis;
KW   nervous system disorder; bacterial infection; viral infection; ss;
KW   fungal infection; ocular disorder; wound healing; tissue regeneration;
KW   epithelial cell proliferation; skin ageing; chemotaxis; IgG Fc region.
XX
```

```
OS   Homo sapiens.
XX
PN   WO200123547-A1.
XX
PD   05-APR-2001.
XX
PF   26-SEP-2000; 2000WO-US26337.
XX
PR   27-SEP-1999;   99US-0155806.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Komatsoulis GA,  Ruben SM,  Rosen CA;
XX
DR   WPI; 2001-266151/27.
DR   P-PSDB; AAU01575.
XX
PT   Nucleic acids encoding 26 human secreted polypeptides, useful for
PT   preventing, diagnosing and/or treating e.g. Gaucher's disease,
PT   Alzheimer's disease, Scimitar syndrome, Creutzfeldt-Jacob disease,
     diabetes mellitus and multiple sclerosis -
XX
PS   Disclosure; Page 374; 412pp; English.
XX
CC   Sequences AAS02631-AAS02665 represent isolated nucleic acid molecules
CC   and PCR primers of the invention. Secreted proteins and their related
CC   nucleic acids can be used in the diagnosis of or susceptibility to a
CC   pathological condition by determining the presence or absence of a
CC   mutation in a nucleic acid or the presence or amount of expression of a
CC   secreted protein. The sequences are used to prevent, treat or ameliorate
CC   a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
CC   dogs, chickens or sheep. The antibodies to the polypeptides can also be
CC   used in alleviating symptoms associated with disorders and in
CC   diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC   immunosorbent assays (ELISA). The disorders include autoimmune diseases
CC   e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of
CC   the breast or liver, cardiovascular disorders e.g. cardiac arrest,
CC   cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous
CC   system disorders e.g. Alzheimer's disease, infections caused by bacteria,
CC   viruses and fungi and ocular disorders e.g. corneal infection. The
CC   peptides can also be used to aid wound healing and epithelial cell
CC   proliferation, to help prevent skin ageing due to sunburn, to maintain
CC   organs before transplantation, to regenerate tissues, in chemotaxis and
CC   as a food additive or preservative to alter storage capabilities.
XX
SQ   Sequence 1385 BP; 296 A; 397 C; 408 G; 280 T; 4 other;

  Query Match             73.3%;  Score 17.6;  DB 22;  Length 1385;
  Best Local Similarity   83.3%;  Pred. No. 97;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 CTTCTCTATTGTCACCGTGGTCCA 24
           |||||| ||  |||||||||| |||
Db     586 CTTCTCGATCTTCACCGTGGACCA 609


RESULT 7
AAT06480/c
ID   AAT06480 standard; cDNA; 1638 BP.
XX
AC   AAT06480;
XX
DT   21-AUG-1996  (first entry)
XX
DE   Cystathionine gamma synthase (CS) gene.
XX
KW   Methionine; lysine; aspartokinase; lysC; homoserine dehydrogenase;
KW   cystathionine gamma synthase; chloroplast; transit sequence; seed;
KW   storage protein; animal feed; CS; AK-HDH; ds.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             2..1441
FT                   /*tag= a
FT                   /product= Cystathionine gamma synthase.
XX
PN   WO9531554-A1.
XX
PD   23-NOV-1995.
XX
PF   12-MAY-1995;   95WO-US05545.
XX
PR   13-MAY-1994;   94US-0242408.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Falco SC,  Guida AD,  Locke ME;
XX
DR   WPI; 1996-010939/01.
DR   P-PSDB; AAR85310.
XX
PT   Nucleic acid encoding plant cystathionine gamma synthase - used to
PT   increase the methionine content of seeds for improvement of animal
PT   feeds
XX
PS   Claim 20; Page 51-53; 80pp; English.
XX
CC   Four chimeric genes encoding (1) a plant cystathionine gamma
CC   synthase (CS); (2) a feedback insensitive aspartokinase (lysC),
CC   operably linked to a chloroplast transit sequence; (3) a
CC   bifunctional feedback insensitive aspartokinase homoserine
CC   dehydrogenase (AK-HDH), operably linked to a chloroplast transit
CC   sequence; and (4) a methionine rich storage protein (HSZ); all
CC   being operably linked to plant seed specific regulatory sequences,
CC   are used for increasing the methionine content of the seeds of
CC   plants.  Plants having increased methionine content may be
CC   used to produce improved animal feeds.
XX
SQ   Sequence 1638 BP; 433 A; 395 C; 410 G; 400 T; 0 other;

  Query Match             71.7%;  Score 17.2;  DB 17;  Length 1638;
  Best Local Similarity   86.4%;  Pred. No. 1.6e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CTTCTCTATTGTCACCGTGGTC 22
           |||||||| || | ||||||||
Db     472 CTTCTCTAATGCCTCCGTGGTC 451


RESULT 8
AAQ98751/c
ID   AAQ98751 standard; DNA; 1969 BP.
XX
AC   AAQ98751;
XX
DT   03-JAN-1996  (first entry)
XX
DE   DNA encoding murine soluble epoxide hydrolase.
XX
KW   Epoxide hydrolase; soluble; toxic; carcinogenic; diol;
KW   cis-epoxy-eicosatrienoic acid; vic-hydroxy-eicosatrienoic acid; ss.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1662
FT                   /*tag= a
FT                   /product= Soluble epoxide hydrolase.
XX
PN   US5445956-A.
XX
PD   29-AUG-1995.
XX
PF   13-AUG-1993;   93US-0106761.
XX
```

```
PR   13-AUG-1993;   93US-0106761.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Beetham JK,  Grant DF,  Hammock BD;
XX
DR   WPI; 1995-310896/40.
DR   P-PSDB; AAR80445.
XX
PT   DNA coding for human soluble epoxide hydrolase - used to degrade
PT   potentially toxic or carcinogenic epoxides
XX
PS   Disclosure; Columns 27-32; 21pp; English.
XX
CC   Soluble epoxide hydrolases catalyse the hydrolysis of potentially
CC   toxic or carcinogenic epoxides to the corresponding diols and are
CC   believed to play a role in the formation or degradation of
CC   endogenous chemical mediators e.g. cis-epoxy-eicosatrienoic and vic-
CC   dihydroxy-eicosatrienoic acids.

     Sequence 1969 BP; 533 A; 468 C; 499 G; 469 T; 0 other;

  Query Match             71.7%;  Score 17.2;  DB 16;  Length 1969;
  Best Local Similarity   86.4%;  Pred. No. 1.6e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 TCTCTATTGTCACCGTGGTCCA 24
           ||||| ||||| |||| |||||
Db     401 TCTCTCTTGTCTCCGTCGTCCA 380


RESULT 9
ABK63809/c
ID   ABK63809 standard; cDNA; 1992 BP.
XX
AC   ABK63809;
XX
DT   18-JUN-2002  (first entry)
XX
DE   Rat sequence differentially expressed in response to a hepatotoxin #1716.
XX
KW   Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
KW   differential expression; centrilobular necrosis; steatosis.
XX
OS   Rattus norvegicus.
XX
PN   WO200210453-A2.
XX
     07-FEB-2002.
XX
PF   30-JUL-2001; 2001WO-US23872.
XX
PR   31-JUL-2000; 2000US-222040P.
PR   02-NOV-2000; 2000US-244880P.
PR   11-MAY-2001; 2001US-290029P.
PR   15-MAY-2001; 2001US-290645P.
PR   22-MAY-2001; 2001US-292336P.
PR   06-JUN-2001; 2001US-295798P.
PR   13-JUN-2001; 2001US-297457P.
PR   19-JUN-2001; 2001US-298884P.
PR   09-JUL-2001; 2001US-303459P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Mendrick D,  Porter MW,  Johnson KR,  Castle AL,  Elashoff MR;
XX
DR   WPI; 2002-241625/29.
XX
PT   Predicting toxic effects of compounds or the progression of these toxic
PT   effects by determining the changes in gene expression in tissues or
PT   cells exposed to the toxin and comparing these to gene expression in
PT   unexposed tissues or cells -
XX
PS   Claim 1; Seq ID No 1716; 239pp; English.
XX
CC   The invention relates to methods for predicting toxic effects of
CC   compounds or the progression of these toxic effects by determining the
CC   global changes in gene expression in tissues or cells exposed to the
CC   toxin and comparing these to gene expression in unexposed tissues or
CC   cells. Also included are methods of predicting at least one toxic
CC   effect of a compound or progression of a toxic effect, preferably the
CC   hepatotoxicity of a compound, comprising detecting the level of
CC   expression in a tissue or cell sample exposed to the compound of two or
CC   more genes listed in the specification, where  differential expression of
CC   the genes is indicative of at least one toxic effect or progression.
CC   The method can also be used to identify an agent which modulates the
CC   toxic response and predict cellular pathways that a compound modulates
CC   in a cell. The methods utilise a set of at least two probes (on a solid
CC   support in kit form), where each of the probes comprises a sequence that
CC   specifically hybridises to a gene listed in the specification, a computer
CC   system comprising a database containing information identifying the
CC   expression level in a tissue or cell sample exposed to a hepatotoxin of a
CC   set of genes comprising at least two genes listed in the specification,
CC   and a user interface to view the information used to present information
CC   identifying the expression level in a tissue or cell of at least one gene
CC   listed in the specification. The method is useful for elucidating global
CC   changes in gene expression and for identifying toxicity markers in
CC   tissues or cell exposed to a known toxin. The genes may be used as
CC   toxicity markers in drug screening and toxicity assays. The genes and
CC   gene expression information may be used as diagnostic markers for the
CC   prediction or identification of the physiological state of tissue or cell
CC   sample that has been exposed to a compound or agent. Hepatotoxicity
CC   is characterised by centrilobular necrosis and steatosis. The present
CC   sequence is an expressed sequence tag (EST) or cDNA derived from a gene
CC   which is differentially expressed in response to a hepatotoxic agent.
XX
SQ   Sequence 1992 BP; 517 A; 478 C; 525 G; 472 T; 0 other;

  Query Match             71.7%;  Score 17.2;  DB 24;  Length 1992;
  Best Local Similarity   86.4%;  Pred. No. 1.6e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 TCTCTATTGTCACCGTGGTCCA 24
           ||||| ||||||| || |||||
Db     442 TCTCTCTTGTCACTGTCGTCCA 421


RESULT 10
ABL27042
ID   ABL27042 standard; DNA; 2702 BP.
XX
AC   ABL27042;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 32599.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
```

```
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 32599; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2702 BP; 829 A; 532 C; 524 G; 817 T; 0 other;

  Query Match             71.7%;  Score 17.2;  DB 23;  Length 2702;
  Best Local Similarity   86.4%;  Pred. No. 1.7e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 TTCTCTATTGTCACCGTGGTCC 23
           |||||| |||||||| ||| ||
Db    1400 TTCTCTGTTGTCACCCTGGGCC 1421


RESULT 11
AAT06481/c
ID   AAT06481 standard; cDNA; 3639 BP.
XX
AC   AAT06481;
XX
DT   22-AUG-1996  (first entry)
XX
DE   Cystathionine gamma synthase (CS) XbaI gene fragment.
XX
KW   Methionine; lysine; aspartokinase; lysC; homoserine dehydrogenase;
KW   cystathionine gamma synthase; chloroplast; transit sequence; seed;
KW   storage protein; animal feed; CS; AK-HDH; ss.
XX
OS   Zea mays.
XX
PN   WO9531554-A1.
XX
     23-NOV-1995.

     12-MAY-1995;   95WO-US05545.
XX
PR   13-MAY-1994;   94US-0242408.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Falco SC,  Guida AD,  Locke ME;
XX
DR   WPI; 1996-010939/01.
XX
PT   Nucleic acid encoding plant cystathionine gamma synthase - used to
PT   increase the methionine content of seeds for improvement of animal
PT   feeds
XX
PS   Claim 21; Page 66-68; 80pp; English.
XX
CC   Four chimeric genes encoding (1) a plant cystathionine gamma
CC   synthase (CS); (2) a feedback insensitive aspartokinase (lysC),
CC   operably linked to a chloroplast transit sequence; (3) a
CC   bifunctional feedback insensitive aspartokinase homoserine
CC   dehydrogenase (AK-HDH), operably linked to a chloroplast transit
CC   sequence; and (4) a methionine rich storage protein (HSZ); all
CC   being operably linked to plant seed specific regulatory sequences,
CC   are used for increasing the methionine content of the seeds of
CC   plants.  Plants having increased methionine content may be
CC   used to produce improved animal feeds.
XX
SQ   Sequence 3639 BP; 957 A; 806 C; 781 G; 1095 T; 0 other;

  Query Match             71.7%;  Score 17.2;  DB 17;  Length 3639;
  Best Local Similarity   86.4%;  Pred. No. 1.7e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CTTCTCTATTGTCACCGTGGTC 22
           |||||||| || | ||||||||
Db    1555 CTTCTCTAATGCCTCCGTGGTC 1534


RESULT 12
AAV15589/c
ID   AAV15589 standard; DNA; 261 BP.
XX
AC   AAV15589;
XX
DT   02-JUL-1998  (first entry)
XX
DE   Human HPK-1A C21.7 DNA.
XX
KW   Cervical cancer; treatment; diagnosis; passage cell; lesion;
KW   human foreskin keratinocyte cell line; HPK-1A; antibody; smear; ss.
XX
OS   Homo sapiens.
XX
PN   DE19649207-C1.
XX
PD   26-FEB-1998.
XX
PF   27-NOV-1996;   96DE-1049207.
XX
PR   27-NOV-1996;   96DE-1049207.
XX
PA   (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX
PI   Duerst M,  Nees M;
XX
DR   WPI; 1998-121623/12.
XX
PT   Nucleic acid characteristic of late or early passage cells
PT   immortalised by papilloma virus - and related polypeptide(s) and
PT   antibodies, used for diagnosis and treatment of cervical cancer and
PT   assessing potential for progression of cervical lesions
XX
PS   Claim 3; Fig 2; 8pp; German.
XX
CC   This sequence, C21.7, is derived from a human papillomavirus (HPV)
CC   immortalised human foreskin keratinocyte cell line HPK-1A and is
CC   characteristic of late or early passage cells. This sequence is used
CC   in a method for assessing the potential for progression of cervical
CC   lesions. Antibodies generated against the encoded polypeptide are used
CC   for diagnosis of cervical cancer and to assess potential for lesion
CC   progression. Antibodies can also be used therapeutically by inhibiting
CC   the polypeptide. Antisense molecules based on the nucleotide sequence
CC   are used to inhibit expression of the protein. Detecting polypeptides,
CC   or related RNA, characteristic of late passage cells (which are
CC   potentially malignant) in cervical smears is a reliable way of
CC   assessing progression potential.
XX
SQ   Sequence 261 BP; 69 A; 56 C; 77 G; 59 T; 0 other;

  Query Match             70.0%;  Score 16.8;  DB 19;  Length 261;
  Best Local Similarity   90.0%;  Pred. No. 1.8e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 CTCTATTGTCACCGTGGTCC 23
           |||||||||||||| | ||||
Db      31 CTCTATTGTCACCCTCGTCC 12
```

```
RESULT 13
AAF66319
ID   AAF66319 standard; cDNA; 400 BP.
XX
AC   AAF66319;
XX
DT   09-APR-2001  (first entry)
XX
DE   Novel human polynucleotide, SEQ ID NO: 2075.
XX
KW   Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW   breast cancer; lung cancer; cancer detection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200102568-A2.
XX
     11-JAN-2001.

PF   30-JUN-2000; 2000WO-US18374.
XX
PR   02-JUL-1999;   99US-0142310.
PR   02-JUL-1999;   99US-0142311.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI   Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI   Crkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitiz D;
PI   Kita D, Garcia V, Jones LW, Strache-Crain B;
XX
DR   WPI; 2001-091805/10.
XX
PT   Library of polynucleotides for diagnosing a cancerous state of a
PT   mammalian cell and detecting cancer, particularly of the colon or
PT   prostate, comprises 3351 human polynucleotide sequences -
XX
PS   Claim 9; Page 845; 1046pp; English.
XX
CC   The present sequence is one of 3351 sequences in a library of human
CC   polynucleotides. The library is used to detect differentially expressed
CC   genes correlated with a cancerous state of a mammalian cell and can
CC   detect colon, prostate, breast and lung cancer. The library can be used
CC   to produce probes for detection of mRNA and to produce additional copies
CC   of the polynucleotides. The probes can be used for chromosome mapping of
CC   the polynucleotide and for detection of transcription levels. Ribozymes
CC   or antisense oligonucleotides can be generated. The polynucleotides and
CC   their gene products are used as genetic or biochemical markers (e.g. in
CC   blood or tissues) that will detect the earliest changes along the
CC   carcinogenesis pathway and/or monitor the efficacy of therapies and
CC   preventive interventions. The polynucleotides, polypeptides and
CC   antibodies against them can be used in pharmaceutical compositions to
CC   treat the cancers and proliferative disorders such as neoplasia,
CC   dysplasia and hyperplasia.
XX
SQ   Sequence 400 BP; 90 A; 111 C; 111 G; 87 T; 1 other;

  Query Match             70.0%;  Score 16.8;  DB 22;  Length 400;
  Best Local Similarity   90.0%;  Pred. No. 2e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CTTCTCTATTGTCACCGTGG 20
           || ||||||||||||||| |
Db     374 CTCCTCTATTGTCACCGTTG 393


RESULT 14
ABQ55843/c
ID   ABQ55843 standard; cDNA; 502 BP.
XX
AC   ABQ55843;
XX
DT   22-AUG-2002  (first entry)
XX
DE   Human ovarian antigen HOVKE66 cDNA, SEQ ID NO:1723.
XX
KW   Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW   ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW   infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW   PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW   inflammatory condition; immune disorder; blood disorder;
KW   cardiovascular disorder; respiratory disorder; neurological disorder;
KW   gastrointestinal disorder; urinary system disorder; drug screening;
KW   gene therapy; chromosome mapping; forensic analysis;
KW   antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW   antiinflammatory; gynaecological; reproductive; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200200677-A1.
XX
PD   03-JAN-2002.
XX
PF   07-JUN-2001; 2001WO-US18569.
XX
PR   07-JUN-2000; 2000US-209467P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Birse CE,  Rosen CA;
XX
DR   WPI; 2002-147878/19.
DR   P-PSDB; ABP42766.
XX
PT   Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT   useful in the prevention, treatment and diagnosis of cancer (e.g.
PT   ovarian cancer), immune disorders, cardiovascular disorders and
PT   neurological diseases -
XX
PS   Claim 1; SEQ ID No 1723; 2922pp; English.
XX
CC   The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC   ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC   encompasses polypeptides 90% identical and polynucleotides 95% identical
CC   to the sequences of the invention. The invention additionally relates to
CC   recombinant vectors and host cells comprising human ovarian antigen
CC   polynucleotides, antibodies against human ovarian antigens, and the use
CC   of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC   treating, prognosing or preventing various ovary and/or breast-related
CC   disorders. Such conditions include ovarian cancer and breast cancer, and
CC   metastatic tumours of ovarian or breast origin, reproductive system
CC   disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC   polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC   disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC   shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC   vaginitis), immune disorders (e.g., congenital and acquired
CC   immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC   blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC   respiratory disorders, neurological disorders, gastrointestinal disorders
CC   and urinary system disorders. Ovarian antigen polypeptides and
CC   polynucleotides may also be used in screening for compounds which
CC   modulate ovarian antigen expression or activity. The polynucleotides may
CC   further be used for gene therapy, chromosome mapping, in the
CC   identification of individuals and in forensic analysis, and the
CC   polypeptides may be used as food additives or to prepare antibodies
CC   useful in disease diagnosis, drug targeting and phenotyping. The present
CC   sequence represents cDNA encoding a human ovarian antigen of the
CC   invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 502 BP; 112 A; 125 C; 152 G; 113 T; 0 other;
```

```
  Query Match             70.0%;  Score 16.8;  DB 24;  Length 502;
  Best Local Similarity   90.0%;  Pred. No. 2e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 CTCTATTGTCACCGTGGTCC 23
           |||||||||||| | ||||
Db     113 CTCTATTGTCACCCTCGTCC 94


RESULT 15
AAS33242
ID   AAS33242 standard; cDNA; 904 BP.
XX
AC   AAS33242;
XX
DT   04-DEC-2001  (first entry)
XX
DE   DNA encoding human secreted protein, Seq ID No 201.
XX
KW   Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
KW   rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
KW   cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
KW   cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
KW   multiple sclerosis; cancer; hyperproliferative disorder; infection;
KW   Gaucher's disease; neurological disease; cerebrovascular disorder;
KW   thrombosis; wound healing; ss.
XX
OS   Homo sapiens.
XX
PN   WO200155326-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01347.
XX
PR   31-JAN-2000; 2000US-0179065.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-451931/48.
DR   P-PSDB; AAU20533.
XX
PT   New nucleic acids and polypeptides, useful for diagnosing, preventing
PT   or treating medical conditions -
XX
PS   Claim 1; SEQ ID No 201; 753pp; English.

CC   The invention relates to novel isolated nucleic acid molecules (I)
CC   encoding human secreted proteins (II). (I) and (II) are used to prevent,
CC   treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
CC   goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
CC   the prevention, treatment and diagnosis of diseases associated with
CC   inappropriate expression of secreted proteins. (I) and complementary
CC   sequences may also be used as DNA probes in diagnostic assays (e.g.
CC   polymerase chain reactions (PCR)) to detect and quantitate the presence
CC   of similar nucleic acid sequences in samples, and so which patients may
CC   be in need of restorative therapy. (II) may also be used as antigens in
CC   the production of antibodies and in assays to identify modulators
CC   (agonists and antagonists) of the expression and activity of the secreted
CC   proteins. The anti-(II) antibodies and antagonists may also be used to
CC   down regulate expression and activity of (II). The anti-(II) antibodies
CC   may also be used as diagnostic agents for detecting the presence of (II)
CC   in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The
CC   disorders include for example: immune/autoimmune diseases (e.g. HIV
CC   (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC   and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC   melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC   Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC   Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
CC   cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
CC   angina and thrombosis), infections caused by bacteria, viruses and
CC   fungi and ocular disorders (e.g. corneal infections). (I) and (II),
CC   agonists, antagonists and antibodies can also be used to promote wound
CC   healing, maintain organs before transplantation, and support cell culture
CC   of primary tissues. AAS33043-AAS33486 represent human secreted protein
CC   coding sequences, PCR primers, and related sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification but was obtained in electronic format directly from WIPO
CC   at: ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 904 BP; 212 A; 249 C; 262 G; 180 T; 1 other;

  Query Match             70.0%;  Score 16.8;  DB 22;  Length 904;
  Best Local Similarity   90.0%;  Pred. No. 2.2e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CTTCTCTATTGTCACCGTGG 20
           || ||||||||||||||| |
Db     541 CTCCTCTATTGTCACCGTTG 560


Search completed: January 23, 2003, 13:54:56
Job time : 126.667 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 23, 2003, 10:24:00 ; Search time 921.333 Seconds
(without alignments)
758.105 Million cell updates/sec

Title: US-09-700-148-17
Perfect score: 24
Sequence: 1 ggttcctttgacggtgcgatgaag 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		24	100.0	24	6	AX010438	AX010438 Sequence
2		24	100.0	310	6	AX010425	AX010425 Sequence
3		24	100.0	1950	1	SEU43237	U43237 Salmonella
4		24	100.0	1950	1	SEU43238	U43238 Salmonella
5		24	100.0	1950	1	SEU43239	U43239 Salmonella
6		24	100.0	1950	1	SEU43242	U43242 Salmonella
7		24	100.0	1950	1	SEU43246	U43246 Salmonella
8		24	100.0	1950	1	SEU43247	U43247 Salmonella
9		24	100.0	1950	1	SEU43248	U43248 Salmonella
10		24	100.0	1950	1	SEU43249	U43249 Salmonella
11		24	100.0	1950	1	SEU43250	U43250 Salmonella
12		24	100.0	1950	1	SEU43251	U43251 Salmonella
13		24	100.0	1950	1	SEU43252	U43252 Salmonella
14		24	100.0	1950	1	SEU43271	U43271 Salmonella
15		24	100.0	1950	1	SEU43272	U43272 Salmonella
16		24	100.0	1950	1	SGU43273	U43273 Salmonella
17		24	100.0	2176	1	STYINVA	M90846 Salmonella
18	c	24	100.0	23125	1	AE008832	AE008832 Salmonell
19	c	24	100.0	274050	1	AL627276	AL627276 Salmonell
20		22.4	93.3	1950	1	SEU43240	U43240 Salmonella
21		22.4	93.3	1950	1	SEU43241	U43241 Salmonella
22		22.4	93.3	1950	1	SEU43243	U43243 Salmonella
23		22.4	93.3	1950	1	SEU43244	U43244 Salmonella
24		22.4	93.3	1950	1	SEU43245	U43245 Salmonella
25	c	21.4	89.2	10875	1	AE005515	AE005515 Escherich
26	c	21.4	89.2	16950	6	AR204270	AR204270 Sequence
27		18.2	75.8	15343	1	AY053505	AY053505 Bacteroid
28		18.2	75.8	63686	2	AC100396	AC100396 Mus muscu
29		18.2	75.8	79691	9	AC079749	AC079749 Homo sapi
30	c	18.2	75.8	156929	2	AC073751	AC073751 Mus muscu
31		18.2	75.8	178275	2	AC102167	AC102167 Mus muscu
32		17.8	74.2	1898	1	ECU18785	U18785 Escherichia
33	c	17.8	74.2	10314	1	AE005498	AE005498 Escherich
34	c	17.8	74.2	10404	1	AE000354	AE000354 Escherich
35	c	17.8	74.2	10404	6	AX370274	AX370274 Sequence
36	c	17.8	74.2	18484	1	D90892	D90892 E.coli geno
37		17.8	74.2	69260	2	AC017374	AC017374 Drosophil
38	c	17.8	74.2	82700	2	AC111596	AC111596 Rattus no
39	c	17.8	74.2	115530	2	AC120980	AC120980 Rattus no
40		17.8	74.2	126355	2	AC126143	AC126143 Rattus no
41		17.8	74.2	138647	2	AC128259	AC128259 Rattus no
42		17.8	74.2	141908	2	AC123282	AC123282 Rattus no
43		17.8	74.2	153448	2	AC112088	AC112088 Rattus no
44		17.8	74.2	158093	2	AC124888	AC124888 Rattus no
45	c	17.8	74.2	158336	2	AC110336	AC110336 Rattus no

ALIGNMENTS

RESULT 1
AX010438
LOCUS AX010438 24 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 17 from Patent WO9958713.
ACCESSION AX010438
VERSION AX010438.1 GI:9997281
KEYWORDS .
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 24)
AUTHORS Grohmann,L., Gerbling,K.P. and Lauter,F.R.
TITLE Method for detecting microorganisms in products
JOURNAL Patent: WO 9958713-A 17 18-NOV-1999;
GROHMANN LUTZ (DE); BIOINSIDE GMBH (DE); GERBLING KLAUS PETER (DE);

```
            LAUTER FRANK ROMAN (DE)
FEATURES             Location/Qualifiers
     source          1. .24
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Reverse Primer Sequence (#542)"
BASE COUNT        4 a      4 c      9 g      7 t
ORIGIN

  Query Match             100.0%;  Score 24;  DB 6;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 0.13;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GGTTCCTTTGACGGTGCGATGAAG 24
           ||||||||||||||||||||||||
Db       1 GGTTCCTTTGACGGTGCGATGAAG 24


RESULT 2
AX010425
LOCUS       AX010425                 310 bp    DNA     linear   PAT 06-SEP-2000
DEFINITION  Sequence 4 from Patent WO9958713.
ACCESSION   AX010425
VERSION     AX010425.1  GI:9997268
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 310)
  AUTHORS   Grohmann,L., Gerbling,K.P. and Lauter,F.R.
  TITLE     Method for detecting microorganisms in products
  JOURNAL   Patent: WO 9958713-A 4 18-NOV-1999;
            GROHMANN LUTZ (DE); BIOINSIDE GMBH (DE); GERBLING KLAUS PETER (DE);
            LAUTER FRANK ROMAN (DE)
FEATURES             Location/Qualifiers
     source          1. .310
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Primer-Sonde-Primer"
BASE COUNT       64 a     63 c     94 g     89 t
ORIGIN

  Query Match             100.0%;  Score 24;  DB 6;  Length 310;
  Best Local Similarity   100.0%;  Pred. No. 0.13;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GGTTCCTTTGACGGTGCGATGAAG 24
           ||||||||||||||||||||||||
Db     282 GGTTCCTTTGACGGTGCGATGAAG 305


RESULT 3
SEU43237
LOCUS       SEU43237                1950 bp    DNA     linear   BCT 21-MAR-1997
DEFINITION  Salmonella enterica invasion protein (invA) gene, partial cds.
ACCESSION   U43237
VERSION     U43237.1  GI:1236804
KEYWORDS    .
SOURCE      Salmonella enterica.
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.S., Whittam,T.S. and Selander,R.K.
  TITLE     Molecular genetic relationships of the salmonellae
  JOURNAL   Appl. Environ. Microbiol. 62 (3), 804-808 (1996)
  MEDLINE   97076912
   PUBMED   8975610
REFERENCE   2  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Li,J., Ochman,H. and Selander,R.K.
  TITLE     Comparative genetics of the inv-spa invasion gene complex of
            Salmonella enterica
  JOURNAL   J. Bacteriol. 179 (6), 1985-1991 (1997)
  MEDLINE   97221599
   PUBMED   9068645
REFERENCE   3  (bases 1 to 1950)
  AUTHORS   Boyd,E.F.-S., Whittam,T.S. and Selander,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) E. Fidelma Boyd, OEB, Harvard University,
            16 Divinity Ave., Cambridge, MA 02138, USA
FEATURES             Location/Qualifiers
     source          1. .1950
                     /organism="Salmonella enterica"
                     /strain="RKS4194"
                     /db_xref="taxon:28901"
                     /map="59 Min"
     gene            1. .1950
                     /gene="invA"
     CDS             <1. .>1950
                     /gene="invA"
                     /note="InvA; membrane spanning protein"
                     /codon_start=1
                     /transl_table=11
                     /product="invasion protein"
                     /protein_id="AAC45021.1"
                     /db_xref="GI:1236805"
                     /translation="NSARLRPELLILVLMVMIISMFVIPLPTYLVDFLIALNIVLAIL
                     VFMGSFYIDRILSFSTFPAVLLITTLFRLALSISTSRLILIEADAGEIIATFGQFVIG
                     DSLAVGFVVFSIVTVVQFIVITKGSERVAEVAARFSLDGMPGKQMSIDADLKAGIIDA
                     DAARERRSVLERESQLYGSFDGAMKFIKGDAIAGIIIIFVNFIGGISVGMTRHGMDLS
                     SALSTYTMLTIGDGLVAQIPALLIAISAGFIVTRVNGDTDNMGRNIMTQLLNNPFVLV
                     VTAILTISMGTLPGFPLPVFVILSVVLSVLFYFKFREAKRSAAKPKTSKGEQPLSIEE
                     KEGSSLGLIGDLDKVSTETVPLILLVPKSRREDLEKAQLAERLRSQFFIDYGVRLPEV
                     LLRDGEGLDDNSIVLLINEIRVEQFTVYFDLMRVVNYSDEVVSFGINPTIHQQGSSQY
                     FWVTHEEGEKLRELGYVLRNALDELYHCLAVTVARNVNEYFGIQETKHMLDQLEAKFP
                     DLLKEVLRHATVQRISEVLQRLLSERVSVRNMKLIMEALALWAPREKDVINLVEHIRG
                     AMARYICHKFANGGELRAVMVSAEVEDVIRKGIRQTSGSTFLSLDPEASANLMDLITL
                     KLDDLLIAHKDLVLLTSVDVRRFIKK"
BASE COUNT      462 a    399 c    498 g    591 t
ORIGIN

  Query Match             100.0%;  Score 24;  DB 1;  Length 1950;
  Best Local Similarity   100.0%;  Pred. No. 0.13;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GGTTCCTTTGACGGTGCGATGAAG 24
           ||||||||||||||||||||||||
Db     532 GGTTCCTTTGACGGTGCGATGAAG 555


RESULT 4
SEU43238
LOCUS       SEU43238                1950 bp    DNA     linear   BCT 21-MAR-1997
DEFINITION  Salmonella enterica invasion protein (invA) gene, partial cds.
ACCESSION   U43238
VERSION     U43238.1  GI:1236806
KEYWORDS    .
SOURCE      Salmonella enterica.
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.S., Whittam,T.S. and Selander,R.K.
  TITLE     Molecular genetic relationships of the salmonellae
  JOURNAL   Appl. Environ. Microbiol. 62 (3), 804-808 (1996)
  MEDLINE   97076912
   PUBMED   8975610
REFERENCE   2  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Li,J., Ochman,H. and Selander,R.K.
  TITLE     Comparative genetics of the inv-spa invasion gene complex of
            Salmonella enterica
  JOURNAL   J. Bacteriol. 179 (6), 1985-1991 (1997)
  MEDLINE   97221599
   PUBMED   9068645
REFERENCE   3  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.-S., Whittam,T.S. and Selander,R.K.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) E. Fidelma Boyd, OEB, Harvard University,
            16 Divinity Ave., Cambridge, MA 02138, USA
FEATURES             Location/Qualifiers
     source          1. .1950
                     /organism="Salmonella enterica"
                     /strain="RKS3333"
                     /db_xref="taxon:28901"
                     /map="59 Min"
     gene            1. .1950
                     /gene="invA"
     CDS             <1. .>1950
                     /gene="invA"
                     /note="InvA; membrane spanning protein"
                     /codon_start=1
                     /transl_table=11
                     /product="invasion protein"
                     /protein_id="AAC45022.1"
                     /db_xref="GI:1236807"
                     /translation="NSARLRPELLILVLMVMIISMFVIPLPTYLVDFLIALNIVLAIL
                     VFMGSFYIDRILSFSTFPAVLLITTLFRLALSISTSRLILIEADAGEIIATFGQFVIG
                     DSLAVGFVVFSIVTVVQFIVITKGSERVAEVAARFSLDGMPGKQMSIDADLKAGIIDA
                     DAARERRSVLERESQLYGSFDGAMKFIKGDAIAGIIIIFVNFIGGISVGMTRHGMDLS
                     SALSTYTMLTIGDGLVAQIPALLIAISAGFIVTRVNGDTDNMGRNIMTQLLNNPFVLV
                     VTAILTISMGTLPGFPLPVFVILSVVLSVLFYFKFREAKRSAAKPKTSKGEQPLSIEE
                     KEGSSLGLIGDLDKVSTETVPLILLVPKSRREDLEKAQLAERLRSQFFIDYGVRLPEV
                     LLRDGEGLDDNSIVLLINEIRVEQFTVYFDLMRVVNYSDEVVSFGINPTIHQQGSSQY
                     FWVTHEEGEKLRELGYVLRNALDELYHCLAVTVARNVNEYFGIQETKHMLDQLEAKFP
                     DLLKEVLRHATVQRISEVLQRLLSERVSVRNMKLIMEALALWAPREKDVINLVEHIRG
                     AMARYICHKFANGGELRAVMVSAEVEDVIRKGIRQTSGSTFLSLDPEASANLMDLITL
                     KLDDLLIAHKDLVLLTSVDVRRFIKK"
BASE COUNT      463 a    401 c    497 g    589 t
ORIGIN

  Query Match           100.0%;  Score 24;  DB 1;  Length 1950;
  Best Local Similarity  100.0%;  Pred. No. 0.13;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGTTCCTTTGACGGTGCGATGAAG 24
          ||||||||||||||||||||||||
Db    532 GGTTCCTTTGACGGTGCGATGAAG 555


RESULT 5
SEU43239
LOCUS       SEU43239                1950 bp    DNA     linear   BCT 21-MAR-1997
DEFINITION  Salmonella enterica invasion protein (invA) gene, partial cds.
  CESSION   U43239
  RSION     U43239.1  GI:1236808
KEYWORDS    .
SOURCE      Salmonella enterica.
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.S., Whittam,T.S. and Selander,R.K.
  TITLE     Molecular genetic relationships of the salmonellae
  JOURNAL   Appl. Environ. Microbiol. 62 (3), 804-808 (1996)
  MEDLINE   97076912
   PUBMED   8975610
REFERENCE   2  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Li,J., Ochman,H. and Selander,R.K.
  TITLE     Comparative genetics of the inv-spa invasion gene complex of
            Salmonella enterica
  JOURNAL   J. Bacteriol. 179 (6), 1985-1991 (1997)
  MEDLINE   97221599
   PUBMED   9068645
REFERENCE   3  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.-S., Whittam,T.S. and Selander,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) E. Fidelma Boyd, OEB, Harvard University,
            16 Divinity Ave., Cambridge, MA 02138, USA
FEATURES             Location/Qualifiers
     source          1. .1950
                     /organism="Salmonella enterica"
                     /strain="RKS3057"
                     /db_xref="taxon:28901"
                     /map="59 Min"
     gene            1. .1950
                     /gene="invA"
     CDS             <1. .>1950
                     /gene="invA"
                     /note="InvA; membrane spanning protein"
                     /codon_start=1
                     /transl_table=11
                     /product="invasion protein"
                     /protein_id="AAC45023.1"
                     /db_xref="GI:1236809"
                     /translation="NSARLRPELLILVLMVMIISMFVIPLPTYLVDFLIALNIVLAIL
                     VFMGSFYIDRILSFSTFPAVLLITTLFRLALSISTSRLILIEADAGEIIATFGQFVIG
                     DSLAVGFVVFSIVTVVQFIVITKGSERVAEVAARFSLDGMPGKQMSIDADLKAGIIDA
                     DAARERRSVLERESQLYGSFDGAMKFIKGDAIAGIIIIFVNFIGGISVGMTRHGMDLS
                     SALSTYTMLTIGDGLVAQIPALLIAISAGFIVTRVNGDTDNMGRNIMTQLLNNPFVLV
                     VTAILTISMGTLPGFPLPVFVILSVVLSVLFYFKFREAKRSAAKPKTSKGEQPLSIEE
                     KEGTSLGLIGDLDKVSTETVPLILLVPKSRREDLEKAQLADRLRSQFFIDYGVRLPEV
                     LLRDGDGLDDNSIILLINEIRVEQFTVYFDLMRVVNYSDEVASFGINPTTHQQGSSQY
                     FWVTHEEGEKLRELGYVLRNALDELYHCLAVTVARNVNEYFGIQETKHMLDQLEAKFP
                     DLLKEVLRHATVQRISEVLQRLLSERVSVRNMKLIMEALALWAPREKDVINLVEHIRG
                     AMARYICHKFANGGELRAVMVSAEVEDVIRKGIRQTSGSTFLSLDPEASANLMDLITL
                     KLDDLLIAHKDLVLLTSVDVRRFIKK"
BASE COUNT      464 a    414 c    493 g    579 t
ORIGIN

  Query Match           100.0%;  Score 24;  DB 1;  Length 1950;
  Best Local Similarity  100.0%;  Pred. No. 0.13;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGTTCCTTTGACGGTGCGATGAAG 24
          ||||||||||||||||||||||||
Db    532 GGTTCCTTTGACGGTGCGATGAAG 555


RESULT 6
SEU43242
LOCUS       SEU43242                1950 bp    DNA     linear   BCT 21-MAR-1997
DEFINITION  Salmonella enterica invasion protein (invA) gene, partial cds.
ACCESSION   U43242
VERSION     U43242.1  GI:1236814
KEYWORDS    .
SOURCE      Salmonella enterica.
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.S., Whittam,T.S. and Selander,R.K.
  TITLE     Molecular genetic relationships of the salmonellae
  JOURNAL   Appl. Environ. Microbiol. 62 (3), 804-808 (1996)
  MEDLINE   97076912
   PUBMED   8975610
REFERENCE   2  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Li,J., Ochman,H. and Selander,R.K.
  TITLE     Comparative genetics of the inv-spa invasion gene complex of
            Salmonella enterica
  JOURNAL   J. Bacteriol. 179 (6), 1985-1991 (1997)
  MEDLINE   97221599
   PUBMED   9068645
REFERENCE   3  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.-S., Whittam,T.S. and Selander,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) E. Fidelma Boyd, OEB, Harvard University,
            16 Divinity Ave., Cambridge, MA 02138, USA
FEATURES             Location/Qualifiers
     source          1. .1950
                     /organism="Salmonella enterica"
                     /strain="RKS3027"
                     /db_xref="taxon:28901"
```

```
                     /map="59 Min"
     gene            1. .1950
                     /gene="invA"
     CDS             <1. .>1950
                     /gene="invA"
                     /note="InvA; membrane spanning protein"
                     /codon_start=1
                     /transl_table=11
                     /product="invasion protein"
                     /protein_id="AAC45026.1"
                     /db_xref="GI:1236815"
                     /translation="NSARLRPELLILVLMVMIISMFVIPLPTYLVDFLIALNIVLAIL
                     VFMGSFYIDRILSFSTFPAVLLITTLFRLALSISTSRLILIEADAGEIIATFGQFVIG
                     DSLAVGFVVFSIVTVVQFIVITKGSERVAEVAARFSLDGMPGKQMSIDADLKAGIIDA
                     DAARERRSVLERESQLYGSFDGAMKFIKGDAIAGIIIIFVNFIGGISVGMTRHGMDLS
                     SALSTYTMLTIGDGLVAQIPALLIAISAGFIVTRVNGDTDNMGRNIMTQLLNNPFVLV
                     VTAILTISMGTLPGFPLPVFVILAVVLSVLFYFKFREAKRSAAKPKTSKGEQPLSIEE
                     KEGTSLGLIGDLDKVSTETVPLILLVPKSRREDLEKAQLADRLRSQFFIDYGVRLPEV
                     LLRDGEGVSDNSIVLLINEIRVEQFTVYFDLMRVVNYSDEVASFGINPTTYHQGSSQY
                     FWVTHEEGEKLRELGYVLRNALDELYHCLAVTLARNVNEYFGIQETKHMLDQLEAKFP
                     DLLKEVLRHATVQRISEVLQRLLSERVSVRNMKLIMEALALWAPREKDVINLVEHIRG
                     AMARYICHKFANGGELRAVMVSAEVEDVIRKGIRQTSGSTFLSLEPEASANLMDLITL
                     KLDDLLIAHKDLVLLTSVDVRRFIKK"
BASE COUNT      465 a    398 c    496 g    591 t
ORIGIN

  Query Match             100.0%;  Score 24;  DB 1;  Length 1950;
  Best Local Similarity   100.0%;  Pred. No. 0.13;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GGTTCCTTTGACGGTGCGATGAAG 24
           ||||||||||||||||||||||||
Db     532 GGTTCCTTTGACGGTGCGATGAAG 555


RESULT 7
SEU43246
LOCUS       SEU43246                1950 bp    DNA     linear   BCT 21-MAR-1997
DEFINITION  Salmonella enterica invasion protein (invA) gene, partial cds.
ACCESSION   U43246
VERSION     U43246.1  GI:1236822
KEYWORDS    .
SOURCE      Salmonella enterica.
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.S., Whittam,T.S. and Selander,R.K.
  TITLE     Molecular genetic relationships of the salmonellae
  JOURNAL   Appl. Environ. Microbiol. 62 (3), 804-808 (1996)
  MEDLINE   97076912
   PUBMED   8975610
REFERENCE   2  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Li,J., Ochman,H. and Selander,R.K.
  TITLE     Comparative genetics of the inv-spa invasion gene complex of
            Salmonella enterica
  JOURNAL   J. Bacteriol. 179 (6), 1985-1991 (1997)
  MEDLINE   97221599
   PUBMED   9068645
REFERENCE   3  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.-S., Whittam,T.S. and Selander,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) E. Fidelma Boyd, OEB, Harvard University,
            16 Divinity Ave., Cambridge, MA 02138, USA
FEATURES             Location/Qualifiers
     source          1. .1950
                     /organism="Salmonella enterica"
                     /strain="RKS2995"
                     /db_xref="taxon:28901"
                     /map="59 Min"
     gene            1. .1950
                     /gene="invA"
     CDS             <1. .>1950
                     /gene="invA"
                     /note="InvA; membrane spanning protein"
                     /codon_start=1
                     /transl_table=11
                     /product="invasion protein"
                     /protein_id="AAC45030.1"
                     /db_xref="GI:1236823"
                     /translation="NSARLRPELLILVLMVMIISMFVIPLPTYLVDFLIALNIVLAIL
                     VFMGSFYIDRILSFSTFPAVLLITTLFRLALSISTSRLILIEADAGEIIATFGQFVIG
                     DSLAVGFVVFSIVTVVQFIVITKGSERVAEVAARFSLDGMPGKQMSIDADLKAGIIDA
                     DAARERRSVLERESQLYGSFDGAMKFIKGDAIAGIIIIFVNFIGGISVGMTRHGMDLS
                     SALSTYTMLTIGDGLVAQIPALLIAISAGFIVTRVNGDTDNMGRNIMTQLLNNPFVLV
                     VTAILTISMGTLPGFPLPVFVILSVVLSVLFYFKFREAKRSAAKPKTSKGEQPLSIEE
                     KEGTSLGLIGDLDKVSTETVPLILLVPKSRREDLEKAQLADRLRSQFFIDYGVRLPEV
                     LLRDGDGLDDNSIILLINEIRVEQFTVYFDLMRVVNYSDEVASFGINPTTHQQGSSQY
                     FWVTHEEGEKLRELGYVLRNALDELYHCLAVTVARNVNEYFGIQETKHMLDQLEAKFP
                     DLLKEVLRHATVQRISEVLQRLLSERVSVRNMKLIMEALALWAPREKDVINLVEHIRG
                     AMARYICHKFANGGELRAVMVSAEVEDVIRKGIRQTSGSTFLSLDPEASANLMDLITL
                     KLDDLLIAHKDLVLLTSVDVRRFIKK"
BASE COUNT      466 a    409 c    494 g    581 t
ORIGIN

  Query Match             100.0%;  Score 24;  DB 1;  Length 1950;
  Best Local Similarity   100.0%;  Pred. No. 0.13;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GGTTCCTTTGACGGTGCGATGAAG 24
           ||||||||||||||||||||||||
Db     532 GGTTCCTTTGACGGTGCGATGAAG 555


RESULT 8
SEU43247
LOCUS       SEU43247                1950 bp    DNA     linear   BCT 21-MAR-1997
DEFINITION  Salmonella enterica invasion protein (invA) gene, partial cds.
ACCESSION   U43247
VERSION     U43247.1  GI:1236824
KEYWORDS    .
SOURCE      Salmonella enterica.
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.S., Whittam,T.S. and Selander,R.K.
  TITLE     Molecular genetic relationships of the salmonellae
  JOURNAL   Appl. Environ. Microbiol. 62 (3), 804-808 (1996)
  MEDLINE   97076912
   PUBMED   8975610
REFERENCE   2  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Li,J., Ochman,H. and Selander,R.K.
  TITLE     Comparative genetics of the inv-spa invasion gene complex of
            Salmonella enterica
  JOURNAL   J. Bacteriol. 179 (6), 1985-1991 (1997)
  MEDLINE   97221599
   PUBMED   9068645
REFERENCE   3  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.-S., Whittam,T.S. and Selander,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) E. Fidelma Boyd, OEB, Harvard University,
            16 Divinity Ave., Cambridge, MA 02138, USA
FEATURES             Location/Qualifiers
     source          1. .1950
                     /organism="Salmonella enterica"
                     /strain="RKS2993"
                     /db_xref="taxon:28901"
                     /map="59 Min"
     gene            1. .1950
                     /gene="invA"
     CDS             <1. .>1950
                     /gene="invA"
                     /note="InvA; membrane spanning protein"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="invasion protein"
                     /protein_id="AAC45031.1"
                     /db_xref="GI:1236825"
                     /translation="NSARLRPELLILVLMVMIISMFVIPLPTYLVDFLIALNIVLAIL
                     VFMGSFYIDRILSFSTFPAVLLITTLFRLALSISTSRLILIEADAGEIIATFGQFVIG
                     DSLAVGFVVFSIVTVVQFIVITKGSERVAEVAARFSLDGMPGKQMSIDADLKAGIIDA
                     DAARERRSVLERESQLYGSFDGAMKFIKGDAIAGIIIIFVNFIGGISVGMTRHGMDLS
                     SALSTYTMLTIGDGLVAQIPALLIAISAGFIVTRVNGDTDNMGRNIMTQLLNNPFVLV
                     VTAILTISMGTLPGFPLPVFVILSVVLSVLFYFKFREAKRSAAKPKTSKGEQPLSIEE
                     KEGTSLGLIGDLDKVSTETVPLILLVPKSRREDLEKAQLADRLRSQFFIDYGVRLPEV
                     LLRDGEGLDDNSIVLLINEIRVEQFTVYFDLMRVVNYSDEVASFGINPTTHQQGSSQY
                     FWVTHEEGEKLRELGYVLRNALDELYHCLAVTVARNVNEYFGIQETKHMLDQLEAKFP
                     DLLKEVLRHATVQRISEVLQRLLSERVSVRNMKLIMEALALWAPREKDVINLVEHIRG
                     AMARYICHKFANGGELRAVMVSAEVEDIIRKGIRQTSGSTFLSLDPEASANLMDLITL
                     KLDDLLIAHKDLVLLTSVDVRRFIKK"
BASE COUNT      465 a    410 c    493 g    582 t
ORIGIN

  Query Match           100.0%;  Score 24;  DB 1;  Length 1950;
  Best Local Similarity  100.0%;  Pred. No. 0.13;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGTTCCTTTGACGGTGCGATGAAG 24
          ||||||||||||||||||||||||
Db    532 GGTTCCTTTGACGGTGCGATGAAG 555


RESULT 9
SEU43248
LOCUS       SEU43248                1950 bp    DNA     linear   BCT 21-MAR-1997
DEFINITION  Salmonella enterica invasion protein (invA) gene, partial cds.
ACCESSION   U43248
VERSION     U43248.1  GI:1236826
KEYWORDS    .
SOURCE      Salmonella enterica.
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.S., Whittam,T.S. and Selander,R.K.
  TITLE     Molecular genetic relationships of the salmonellae
  JOURNAL   Appl. Environ. Microbiol. 62 (3), 804-808 (1996)
  MEDLINE   97076912
   PUBMED   8975610
REFERENCE   2  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Li,J., Ochman,H. and Selander,R.K.
  TITLE     Comparative genetics of the inv-spa invasion gene complex of
            Salmonella enterica
  JOURNAL   J. Bacteriol. 179 (6), 1985-1991 (1997)
  MEDLINE   97221599
   PUBMED   9068645
REFERENCE   3  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.-S., Whittam,T.S. and Selander,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) E. Fidelma Boyd, OEB, Harvard University,
            16 Divinity Ave., Cambridge, MA 02138, USA
FEATURES             Location/Qualifiers
     source          1. .1950
                     /organism="Salmonella enterica"
                     /strain="RKS2985"
                     /db_xref="taxon:28901"
                     /map="59 Min"
     gene            1. .1950
                     /gene="invA"
     CDS             <1. .>1950
                     /gene="invA"
                     /note="InvA; membrane spanning protein"
                     /codon_start=1
                     /transl_table=11
                     /product="invasion protein"
                     /protein_id="AAC45032.1"
                     /db_xref="GI:1236827"
                     /translation="NSARLRPELLILVLMVMIISMFVIPLPTYLVDFLIALNIVLAIL
                     VFMGSFYIDRILSFSTFPAVLLITTLFRLALSISTSRLILIEADAGEIIATFGQFVIG
                     DSLAVGFVVFSIVTVVQFIVITKGSERVAEVAARFSLDGMPGKQMSIDADLKAGIIDA
                     DAARERRSVLERESQLYGSFDGAMKFIKGDAIAGIIIIFVNFIGGISVGMTRHGMDLS
                     SALSTYTMLTIGDGLVAQIPALLIAISAGFIVTRVNGDTDNMGRNIMTQLLNNPFVLV
                     VTAILTISMGTLPGFPLPVFVILSVVLSVLFYFKFREAKRSAAKPKTSKGEQPLSIEE
                     KEGTSLGLIGDLDKVSTETVPLILLVPKSRREDLEKAQLADRLRSQFFIDYGVRLPEV
                     LLRDGEGLDDNSIVLLINEIRVEQFTVYFDLMRVVNYSDEVASFGINPTTHQQGSSQY
                     FWVTHEEGEKLRELGYVLRNALDELYHCLAVTVARNVNEYFGIQETKHMLDQLEAKFP
                     DLLKEVLRHATVQRISEVLQRLLSERVSVRNMKLIMEALALWAPREKDVINLVEHIRG
                     AMARYICHKFANGGELRAVMVSAEVEDIIRKGIRQTSGSTFLSLDPEASANLMDLITL
                     KLDDLLIAHKDLVLLTSVDVRRFIKK"
BASE COUNT      465 a    411 c    493 g    581 t
ORIGIN

  Query Match           100.0%;  Score 24;  DB 1;  Length 1950;
  Best Local Similarity  100.0%;  Pred. No. 0.13;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGTTCCTTTGACGGTGCGATGAAG 24
          ||||||||||||||||||||||||
Db    532 GGTTCCTTTGACGGTGCGATGAAG 555


RESULT 10
SEU43249
LOCUS       SEU43249                1950 bp    DNA     linear   BCT 21-MAR-1997
DEFINITION  Salmonella enterica invasion protein (invA) gene, partial cds.
ACCESSION   U43249
VERSION     U43249.1  GI:1236828
KEYWORDS    .
SOURCE      Salmonella enterica.
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.S., Whittam,T.S. and Selander,R.K.
  TITLE     Molecular genetic relationships of the salmonellae
  JOURNAL   Appl. Environ. Microbiol. 62 (3), 804-808 (1996)
  MEDLINE   97076912
   PUBMED   8975610
REFERENCE   2  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Li,J., Ochman,H. and Selander,R.K.
  TITLE     Comparative genetics of the inv-spa invasion gene complex of
            Salmonella enterica
  JOURNAL   J. Bacteriol. 179 (6), 1985-1991 (1997)
  MEDLINE   97221599
   PUBMED   9068645
REFERENCE   3  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.-S., Whittam,T.S. and Selander,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) E. Fidelma Boyd, OEB, Harvard University,
            16 Divinity Ave., Cambridge, MA 02138, USA
FEATURES             Location/Qualifiers
     source          1. .1950
                     /organism="Salmonella enterica"
                     /strain="RKS2983"
                     /db_xref="taxon:28901"
                     /map="59 Min"
     gene            1. .1950
                     /gene="invA"
     CDS             <1. .>1950
                     /gene="invA"
                     /note="InvA; membrane spanning protein"
                     /codon_start=1
                     /transl_table=11
                     /product="invasion protein"
                     /protein_id="AAC45033.1"
                     /db_xref="GI:1236829"
                     /translation="NSARLRPELLILVLMVMIISMFVIPLPTYLVDFLIALNIVLAIL
                     VFMGSFYIDRILSFSTFPAVLLITTLFRLALSISTSRLILIEADAGEIIATFGQFVIG
                     DSLAVGFVVFSIVTVVQFIVITKGSERVAEVAARFSLDGMPGKQMSIDADLKAGIIDA
                     DAARERRSVLERESQLYGSFDGAMKFIKGDAIAGIIIIFVNFIGGISVGMTRHGMDLS
                     SALSTYTMLTIGDGLVAQIPALLIAISAGFIVTRVNGDTDNMGRNIMTQLLNNPFVLV
```

```
                    VTAILTISMGTLPGFPLPVFVILSVVLSVLFYFKFREAKRSAGTPKTSKGEQPLSIEE
                    KEGTSLGLIGDLDKVSTETVPLILLVPKSRREDLEKAQLADRLRSQFFIDYGVRLPEV
                    LLRDGEGLDDNSIVLLINEIRVEQFTVYFDLMRVVNYSDEVVSFGINPTTHQQGSSQY
                    FWVTYEEGEKLRELGYVLRNALDELYHCLAVTLARNVNEYFGIQETKHMLDQLEAKFP
                    DLLKEVLRHATVQRISEVLQRLLSERVSVRNMKLIMEALALWAPREKDVINLVEHIRG
                    AMARYICHKFANGGELRAVMVSAEVEDVIRKGIRQTSGSTFLSLEPEASANLMDLITL
                    KLDDLLIAHKDLVLLTSVDVRRFIKK"
BASE COUNT      469 a    399 c    496 g    586 t
ORIGIN

  Query Match           100.0%;  Score 24;  DB 1;  Length 1950;
  Best Local Similarity  100.0%;  Pred. No. 0.13;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGTTCCTTTGACGGTGCCGATGAAG 24
          ||||||||||||||||||||||||
Db    532 GGTTCCTTTGACGGTGCCGATGAAG 555


RESULT 11
SEU43250
LOCUS       SEU43250                1950 bp    DNA     linear   BCT 21-MAR-1997
DEFINITION  Salmonella enterica invasion protein (invA) gene, partial cds.
ACCESSION   U43250
VERSION     U43250.1  GI:1236830
KEYWORDS    .
SOURCE      Salmonella enterica.
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.S., Whittam,T.S. and Selander,R.K.
  TITLE     Molecular genetic relationships of the salmonellae
  JOURNAL   Appl. Environ. Microbiol. 62 (3), 804-808 (1996)
  MEDLINE   97076912
   PUBMED   8975610
REFERENCE   2  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Li,J., Ochman,H. and Selander,R.K.
  TITLE     Comparative genetics of the inv-spa invasion gene complex of
            Salmonella enterica
  JOURNAL   J. Bacteriol. 179 (6), 1985-1991 (1997)
  MEDLINE   97221599
   PUBMED   9068645
REFERENCE   3  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.-S., Whittam,T.S. and Selander,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) E. Fidelma Boyd, OEB, Harvard University,
            16 Divinity Ave., Cambridge, MA 02138, USA
FEATURES             Location/Qualifiers
     source          1. .1950
                     /organism="Salmonella enterica"
                     /strain="RKS2980"
                     /db_xref="taxon:28901"
                     /map="59 Min"
     gene            1. .1950
                     /gene="invA"
     CDS             <1. .>1950
                     /gene="invA"
                     /note="InvA; membrane spanning protein"
                     /codon_start=1
                     /transl_table=11
                     /product="invasion protein"
                     /protein_id="AAC45034.1"
                     /db_xref="GI:1236831"
                     /translation="NSARLRPELLILVLMVMIISMFVIPLPTYLVDFLIALNIVLAIL
                     VFMGSFYIDRILSFSTFPAVLLITTLFRLALSISTSRLILIEADAGEIIATFGQFVIG
                     DSLAVGFVVFSIVTVVQFIVITKGSERVAEVAARFSLDGMPGKQMSIDADLKAGIIDA
                     DAARERRSVLERESQLYGSFDGAMKFIKGDAIAGIIIIFVNFIGGISVGMTRHGMDLS
                     SALSTYTMLTIGDGLVAQIPALLIAISAGFIVTRVNGDTDNMGRNIMTQLLNNPFVLV
                     VTAILTISMGTLPGFPLPVFVILSVVLSVLFYFKFREAKRSAGTPKTSKGEQPLSIEE
                     KEGTSLGLIGDLDKVSTETVPLILLVPKSRREDLEKAQLADRLRSQFFIDYGVRLPEV
                     LLRDGEGLDDNSIVLLINEIRVEQFTVYFDLMRVVNYSDEVVSFGINPTTHQQGSSQY
                     FWVTHEEGEKLRELGYVLRNALDELYHCLAVTLARNVNEYFGIQETKHMLDQLEAKFP
                     DLLKEVLRHATVQRISEVLQRLLSERVSVRNMKLIMEALALWAPREKDVINLVEHIRG
                     AMARYICHKFANGGELRAVMVSAEVEDVIRKGIRQTSGSTFLSLEPEASANLMDLITL
                     KLDDLLIAHKDLVLLTSVDVRRFIKK"
BASE COUNT      469 a    400 c    496 g    585 t
ORIGIN

  Query Match           100.0%;  Score 24;  DB 1;  Length 1950;
  Best Local Similarity  100.0%;  Pred. No. 0.13;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGTTCCTTTGACGGTGCCGATGAAG 24
          ||||||||||||||||||||||||
Db    532 GGTTCCTTTGACGGTGCCGATGAAG 555


RESULT 12
SEU43251
LOCUS       SEU43251                1950 bp    DNA     linear   BCT 21-MAR-1997
DEFINITION  Salmonella enterica invasion protein (invA) gene, partial cds.
ACCESSION   U43251
VERSION     U43251.1  GI:1236832
KEYWORDS    .
SOURCE      Salmonella enterica.
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.S., Whittam,T.S. and Selander,R.K.
  TITLE     Molecular genetic relationships of the salmonellae
  JOURNAL   Appl. Environ. Microbiol. 62 (3), 804-808 (1996)
  MEDLINE   97076912
   PUBMED   8975610
REFERENCE   2  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Li,J., Ochman,H. and Selander,R.K.
  TITLE     Comparative genetics of the inv-spa invasion gene complex of
            Salmonella enterica
  JOURNAL   J. Bacteriol. 179 (6), 1985-1991 (1997)
  MEDLINE   97221599
   PUBMED   9068645
REFERENCE   3  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.-S., Whittam,T.S. and Selander,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) E. Fidelma Boyd, OEB, Harvard University,
            16 Divinity Ave., Cambridge, MA 02138, USA
FEATURES             Location/Qualifiers
     source          1. .1950
                     /organism="Salmonella enterica"
                     /strain="RKS2979"
                     /db_xref="taxon:28901"
                     /map="59 Min"
     gene            1. .1950
                     /gene="invA"
     CDS             <1. .>1950
                     /gene="invA"
                     /note="InvA; membrane spanning protein"
                     /codon_start=1
                     /transl_table=11
                     /product="invasion protein"
                     /protein_id="AAC45035.1"
                     /db_xref="GI:1236833"
                     /translation="NSARLRPELLILVLMVMIISMFVIPLPTYLVDFLIALNIVLAIL
                     VFMGSFYIDRILSFSTFPAVLLITTLFRLALSISTSRLILIEADAGEIIATFGQFVIG
                     DSLAVGFVVFSIVTVVQFIVITKGSERVAEVAARFSLDGMPGKQMSIDADLKAGIIDA
                     DAARERRSVLERESQLYGSFDGAMKFIKGDAIAGIIIIFVNFIGGISVGMTRHGMDLS
                     SALSTYTMLTIGDGLVAQIPALLIAISAGFIVTRVNGDTDNMGRNIMTQLLNNPFVLV
                     VTAILTISMGTLPGFPLPVFVILSVVLSVLFYFKFREAKRSAVKPKTSKGEQPLSIEE
                     KEGTSLGLIGDLDKVSTETVPLILLVPKSRREDLEKAQLADRLRSQFFIDYGVRLPEV
                     LLRDGEGLDDNNIVLLINEIRVEQFTVYFDLMRVINYSDEVASFGINPTTHQQGSSQY
                     FWVTHEEGEKLRELGYVLRNALDELYHCLAVMLARNVNEYFGIQETKYMLDQLEAKFP
                     DLLKEVLRHATVQRISEVLQRLLSERVSVRNMKLIMEALALWAPREKDVINLVEHIRG
                     AMARYICHKFANGGELRAVMVSAEVEDVIRKGIRQTSGSTFLSLDPEASANLMDLITL
                     KLDDLLIAHKDLVLLTSVDVRRFIKK"
BASE COUNT      464 a    399 c    497 g    590 t
```

```
ORIGIN

  Query Match            100.0%;  Score 24;  DB 1;  Length 1950;
  Best Local Similarity  100.0%;  Pred. No. 0.13;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGTTCCTTTGACGGTGCGATGAAG 24
          ||||||||||||||||||||||||
Db    532 GGTTCCTTTGACGGTGCGATGAAG 555


RESULT 13
SEU43252
LOCUS       SEU43252                1950 bp    DNA     linear   BCT 21-MAR-1997
DEFINITION  Salmonella enterica invasion protein (invA) gene, partial cds.
ACCESSION   U43252
VERSION     U43252.1  GI:1236834
KEYWORDS    .
SOURCE      Salmonella enterica.
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.S., Whittam,T.S. and Selander,R.K.
  TITLE     Molecular genetic relationships of the salmonellae
  JOURNAL   Appl. Environ. Microbiol. 62 (3), 804-808 (1996)
  MEDLINE   97076912
   PUBMED   8975610
REFERENCE   2  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Li,J., Ochman,H. and Selander,R.K.
  TITLE     Comparative genetics of the inv-spa invasion gene complex of
            Salmonella enterica
  JOURNAL   J. Bacteriol. 179 (6), 1985-1991 (1997)
  MEDLINE   97221599
   PUBMED   9068645
REFERENCE   3  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.-S., Whittam,T.S. and Selander,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) E. Fidelma Boyd, OEB, Harvard University,
            16 Divinity Ave., Cambridge, MA 02138, USA
FEATURES             Location/Qualifiers
     source          1. .1950
                     /organism="Salmonella enterica"
                     /strain="RKS2978"
                     /db_xref="taxon:28901"
                     /map="59 Min"
     gene            1. .1950
                     /gene="invA"
     CDS             <1. .>1950
                     /gene="invA"
                     /note="InvA; membrane spanning protein"
                     /codon_start=1
                     /transl_table=11
                     /product="invasion protein"
                     /protein_id="AAC45036.1"
                     /db_xref="GI:1236835"
                     /translation="NSARLRPELLILVLMVMIISMFVIPLPTYLVDFLIALNIVLAIL
                     VFMGSFYIDRILSFSTFPAVLLITTLFRLALSISTSRLILIEADAGEIIATFGQFVIG
                     DSLAVGFVVFSIVTVVQFIVITKGSERVAEVAARFSLDGMPGKQMSIDADLKAGIIDA
                     DAARERRSVLERESQLYGSFDGAMKFIKGDAIAGIIIIFVNFIGGISVGMTRHGMDLS
                     SALSTYTMLTIGDGLVAQIPALLIAISAGFIVTRVNGDTDNMGRNIMTQLLNNPFVLV
                     VTAILTISMGTLPGFPLPVFVILSVVLSVLFYFKFREAKRSAVKPKTSKGEQPLSIEE
                     KEGTSLGLIGDLDKVSTETVPLILLVPKSRREDLEKAQLADRLRSQFFIDYGVRLPEV
                     LLRDGEGLDDNNIVLLINEIRVEQFTVYFDLMRVVNYSDEVASFGINPTTHQQGSSQY
                     FWVTHEEGEKLRELGYVLRNALDELYHCLAVMLARNVNEYFGIQETKYMLDQLEAKFP
                     DLLKEVLRHATVQRISEVLQRLLSERVSVRNMKLIMEALALWAPREKDVINLVEHIRG
                     AMARYICHKFANGGELRAVMVSAEVEDVIRKGIRQTSGSTFLSLDPEASANLMDLITL
                     KLDDLLIAHKDLVLLTSVDVRRFIKK"
BASE COUNT      464 a    400 c    498 g    588 t
ORIGIN

  Query Match            100.0%;  Score 24;  DB 1;  Length 1950;
  Best Local Similarity  100.0%;  Pred. No. 0.13;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGTTCCTTTGACGGTGCGATGAAG 24
          ||||||||||||||||||||||||
Db    532 GGTTCCTTTGACGGTGCGATGAAG 555


RESULT 14
SEU43271
LOCUS       SEU43271                1950 bp    DNA     linear   BCT 21-MAR-1997
DEFINITION  Salmonella enterica invasion protein (invA) gene, partial cds.
ACCESSION   U43271
VERSION     U43271.1  GI:1236872
KEYWORDS    .
SOURCE      Salmonella enterica.
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.S., Whittam,T.S. and Selander,R.K.
  TITLE     Molecular genetic relationships of the salmonellae
  JOURNAL   Appl. Environ. Microbiol. 62 (3), 804-808 (1996)
  MEDLINE   97076912
   PUBMED   8975610
REFERENCE   2  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Li,J., Ochman,H. and Selander,R.K.
  TITLE     Comparative genetics of the inv-spa invasion gene complex of
            Salmonella enterica
  JOURNAL   J. Bacteriol. 179 (6), 1985-1991 (1997)
  MEDLINE   97221599
   PUBMED   9068645
REFERENCE   3  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.-S., Whittam,T.S. and Selander,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) E. Fidelma Boyd, OEB, Harvard University,
            16 Divinity Ave., Cambridge, MA 02138, USA
FEATURES             Location/Qualifiers
     source          1. .1950
                     /organism="Salmonella enterica"
                     /strain="RKS1280"
                     /db_xref="taxon:28901"
                     /map="59 Min"
                     /note="serovar Choleraesuis"
     gene            1. .1950
                     /gene="invA"
     CDS             <1. .>1950
                     /gene="invA"
                     /note="InvA; membrane spanning protein"
                     /codon_start=1
                     /transl_table=11
                     /product="invasion protein"
                     /protein_id="AAC45055.1"
                     /db_xref="GI:1236873"
                     /translation="NSARLRPELLILVLMVMIISMFVIPLPTYLVDFLIALNIVLAIL
                     VFMGSFYIDRILSFSTFPAVLLITTLFRLALSISTSRLILIEADAGEIIATFGQFVIG
                     DSLAVGFVVFSIVTVVQFIVITKGSERVAEVAARFSLDGMPGKQMSIDADLKAGIIDA
                     DAARERRSVLERESQLYGSFDGAMKFIKGDAIAGIIIIFVNFIGGISVGMTRHGMDLS
                     SALSTYTMLTIGDGLVAQIPALLIAISAGFIVTRVNGDSDNMGRNIMTQLLNNPFVLV
                     VTAILTISMGTLPGFPLPVFVILSVVLSVLFYFKFREAKRSAAKPKTSKGEQPLSIEE
                     KEGSSLGLIGDLDKVSTETVPLILLVPKSRREDLEKAQLAERLRSQFFIDYGVRLPEV
                     LLRDGEGLDDNSIVLLINEIRVEQFTVYFDLMRVVNYSDEVVSFGINPTIHQQGSSQY
                     FWVTHEEGEKLRELGYVLRNALDELYHCLAVTVARNVNEYFGIQETKHMLDQLEAKFP
                     DLLKEVLRHATVQRISEVLQRLLSERVSVRNMKLIMEALALWAPREKDVINLVEHIRG
                     AMARYICHKFANGGELRAVMVSAEVEDVIRKGIRQTSGSTFLSLDPEASANLMDLITL
                     KLDDLLIAHKDLVLLTSVDVRRFIKK"
BASE COUNT      463 a    397 c    499 g    591 t
ORIGIN

  Query Match            100.0%;  Score 24;  DB 1;  Length 1950;
  Best Local Similarity  100.0%;  Pred. No. 0.13;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGTTCCTTTGACGGTGCGATGAAG 24
```

```
          ||||||||||||||||||||||
Db    532 GGTTCCTTTGACGGTGCGATGAAG 555


RESULT 15
SEU43272
LOCUS       SEU43272                1950 bp    DNA     linear   BCT 21-MAR-1997
DEFINITION  Salmonella enterica invasion protein (invA) gene, partial cds.
ACCESSION   U43272
VERSION     U43272.1  GI:1236874
KEYWORDS    .
SOURCE      Salmonella enterica.
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.S., Whittam,T.S. and Selander,R.K.
  TITLE     Molecular genetic relationships of the salmonellae
  JOURNAL   Appl. Environ. Microbiol. 62 (3), 804-808 (1996)
  MEDLINE   97076912
   PUBMED   8975610
REFERENCE   2  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Li,J., Ochman,H. and Selander,R.K.
  TITLE     Comparative genetics of the inv-spa invasion gene complex of
            Salmonella enterica
  JOURNAL   J. Bacteriol. 179 (6), 1985-1991 (1997)
  MEDLINE   97221599
   PUBMED   9068645
REFERENCE   3  (bases 1 to 1950)
  AUTHORS   Boyd,E.F., Wang,F.-S., Whittam,T.S. and Selander,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) E. Fidelma Boyd, OEB, Harvard University,
            16 Divinity Ave., Cambridge, MA 02138, USA
FEATURES             Location/Qualifiers
     source          1. .1950
                     /organism="Salmonella enterica"
                     /strain="RKS1518"
                     /sub_species="I"
                     /db_xref="taxon:28901"
                     /map="59 Min"
                     /note="serovar Dublin"
     gene            1. .1950
                     /gene="invA"
     CDS             <1. .>1950
                     /gene="invA"
                     /note="InvA; membrane spanning protein"
                     /codon_start=1
                     /transl_table=11
                     /product="invasion protein"
                     /protein_id="AAC45056.1"
                     /db_xref="GI:1236875"
                     /translation="NSARLRPELLILVLMVMIISMFVIPLPTYLVDFLIALNIVLAIL
                     VFMGSFYIDRILSFSTFPAVLLITTLFRLALSISTSRLILIEADAGEIIATFGQFVIG
                     DSLAVGFVVFSIVTVVQFIVITKGSERVAEVAARFSLDGMPGKQMSIDADLKAGIIDA
                     DAARERRSVLERESQLYGSFDGAMKFIKGDAIAGIIIIFVNFIGGISVGMARHGMDLS
                     SALSTYTMLTIGDGLVAQIPALLIAISAGFIVTRVNGDSDNMGRNIMTQLLNNPFVLV
                     VTAILTISMGTLPGFPLPVFVILSVVLSVLFYFKFREAKRSAAKPKTSKGEQPLSIEE
                     KEGSSLGLIGDLDKVSTETVPLILLVPKSRREDLEKAQLAERLRSQFFIDYGVRLPEV
                     LLRDGEGLDDNSIVLLINEIRVEQFTVYFDLMRVVNYSDEVVSFGINPTIHQQGSSQY
                     FWVTHEEGEKLRELGYVLRNALDELYHCLAVTVARNVNEYFGIQETKHMLDQLEAKFP
                     DLLKEVLRHATVQRISEVLQRLLSERVSVRNMKLIMEALALWAPREKDVINLVEHIRG
                     AMARYICHKFANGGELRAVMVSAEVEDVIRKGIRQTSGSTFLSLDPEASANLMDLITL
                     KLDDLLIAHKDLVLLTSVDVRRFIKK"
BASE COUNT      463 a    399 c    498 g    590 t
ORIGIN

  Query Match           100.0%;  Score 24;  DB 1;  Length 1950;
  Best Local Similarity  100.0%;  Pred. No. 0.13;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGTTCCTTTGACGGTGCGATGAAG 24
          ||||||||||||||||||||||||
Db    532 GGTTCCTTTGACGGTGCGATGAAG 555
```

Search completed: January 23, 2003, 13:48:50
Job time : 923.333 secs